

**VOLUME 8, NUMBER 1      JANUARY 2010**

**ISSN:1548-5390 PRINT,1559-176X ONLINE**



**JOURNAL  
OF CONCRETE  
AND APPLICABLE**

**MATHEMATICS**  
**SPECIAL ISSUE I :APPLIED MATHEMATICS**  
**AND APPROXIMATION THEORY**

**EUDOXUS PRESS,LLC**

**SCOPE AND PRICES OF THE JOURNAL**  
**Journal of Concrete and Applicable Mathematics**

A quartely international publication of **Eudoxus Press,LLC**

**Editor in Chief: George Anastassiou**

Department of Mathematical Sciences,  
 University of Memphis  
 Memphis, TN 38152, U.S.A.  
 ganastss@memphis.edu

The main purpose of the "Journal of Concrete and Applicable Mathematics" is to publish high quality original research articles from all subareas of Non-Pure and/or Applicable Mathematics and its many real life applications, as well connections to other areas of Mathematical Sciences, as long as they are presented in a Concrete way. It welcomes also related research survey articles and book reviews. A sample list of connected mathematical areas with this publication includes and is not restricted to: Applied Analysis, Applied Functional Analysis, Probability theory, Stochastic Processes, Approximation Theory, O.D.E, P.D.E, Wavelet, Neural Networks, Difference Equations, Summability, Fractals, Special Functions, Splines, Asymptotic Analysis, Fractional Analysis, Inequalities, Moment Theory, Numerical Functional Analysis, Tomography, Asymptotic Expansions, Fourier Analysis, Applied Harmonic Analysis, Integral Equations, Signal Analysis, Numerical Analysis, Optimization, Operations Research, Linear Programming, Fuzzyness, Mathematical Finance, Stochastic Analysis, Game Theory, Math. Physics aspects, Applied Real and Complex Analysis, Computational Number Theory, Graph Theory, Combinatorics, Computer Science Math. related topics, combinations of the above, etc. In general any kind of Concretely presented Mathematics which is Applicable fits to the scope of this journal. Working Concretely and in Applicable Mathematics has become a main trend in many recent years, so we can understand better and deeper and solve the important problems of our real and scientific world. "Journal of Concrete and Applicable Mathematics" is a peer-reviewed International Quarterly Journal. We are calling for papers for possible publication. The contributor should send three copies of the contribution to the editor in-Chief typed in TEX, LATEX double spaced. [ See: Instructions to Contributors]

**Journal of Concrete and Applicable Mathematics(JCAAM)**

**ISSN:1548-5390 PRINT, 1559-176X ONLINE.**

is published in January, April, July and October of each year by

**EUDOXUS PRESS,LLC,**

1424 Beaver Trail Drive, Cordova, TN38016, USA,

Tel. 001-901-751-3553

anastassioug@yahoo.com

<http://www.EudoxusPress.com>.

**Visit also [www.msci.memphis.edu/~ganastss/jcaam](http://www.msci.memphis.edu/~ganastss/jcaam).**

**Webmaster: Ray Clapsadle**

**Annual Subscription Current Prices:** For USA and Canada, Institutional: Print \$400, Electronic \$250, Print and Electronic \$450. Individual: Print \$150, Electronic

\$80,Print &Electronic \$200.For any other part of the world add \$50 more to the above prices for Print.

Single article PDF file for individual \$15.Single issue in PDF form for individual \$60.

No credit card payments.Only certified check,money order or international check in US dollars are acceptable.

Combination orders of any two from JoCAAA,JCAAM,JAFa receive 25% discount,all three receive 30% discount.

**Copyright**©2010 by Eudoxus Press,LLC all rights reserved.JCAAM is printed in USA.

**JCAAM is reviewed and abstracted by AMS Mathematical Reviews,MATHSCI,and Zentralblatt MATH.**

It is strictly prohibited the reproduction and transmission of any part of JCAAM and in any form and by any means without the written permission of the publisher.It is only allowed to educators to Xerox articles for educational purposes.The publisher assumes no responsibility for the content of published papers.

***JCAAM IS A JOURNAL OF RAPID PUBLICATION***

---

## Editorial Board

### Associate Editors

---

**Editor in -Chief:**

George Anastassiou  
 Department of Mathematical Sciences  
 The University Of Memphis  
 Memphis, TN 38152, USA  
 tel. 901-678-3144, fax 901-678-2480  
 e-mail ganastss@memphis.edu  
[www.msci.memphis.edu/~anastasg/anlyjour.htm](http://www.msci.memphis.edu/~anastasg/anlyjour.htm)  
 Areas: Approximation Theory,  
 Probability, Moments, Wavelet,  
 Neural Networks, Inequalities, Fuzzyness.

**Associate Editors:**

1) Ravi Agarwal  
 Florida Institute of Technology  
 Applied Mathematics Program  
 150 W. University Blvd.  
 Melbourne, FL 32901, USA  
[agarwal@fit.edu](mailto:agarwal@fit.edu)  
 Differential Equations, Difference  
 Equations,  
 Inequalities

2) Drumi D. Bainov  
 Medical University of Sofia  
 P.O. Box 45, 1504 Sofia, Bulgaria  
[drumibainov@yahoo.com](mailto:drumibainov@yahoo.com)  
 Differential Equations, Optimal Control,  
 Numerical Analysis, Approximation Theory

3) Carlo Bardaro  
 Dipartimento di Matematica & Informatica  
 Università di Perugia  
 Via Vanvitelli 1  
 06123 Perugia, ITALY  
 tel. +390755855034, +390755853822,  
 fax +390755855024  
[bardaro@unipg.it](mailto:bardaro@unipg.it) ,  
[bardaro@dipmat.unipg.it](mailto:bardaro@dipmat.unipg.it)  
 Functional Analysis and Approximation Th.,  
 Summability, Signal Analysis, Integral  
 Equations,  
 Measure Th., Real Analysis

4) Francoise Bastin  
 Institute of Mathematics  
 University of Liege  
 4000 Liege

21) Gustavo Alberto Perla Menzala  
 National Laboratory of Scientific Computation  
 LNCC/MCT  
 Av. Getulio Vargas 333  
 25651-075 Petropolis, RJ  
 Caixa Postal 95113, Brasil  
 and

Federal University of Rio de Janeiro  
 Institute of Mathematics  
 RJ, P.O. Box 68530 Rio de Janeiro, Brasil  
[perla@lncc.br](mailto:perla@lncc.br) and [perla@im.ufrj.br](mailto:perla@im.ufrj.br)  
 Phone 55-24-22336068, 55-21-25627513 Ext 224  
 FAX 55-24-22315595  
 Hyperbolic and Parabolic Partial Differential  
 Equations,  
 Exact controllability, Nonlinear Lattices and  
 Global  
 Attractors, Smart Materials

22) Ram N. Mohapatra  
 Department of Mathematics  
 University of Central Florida  
 Orlando, FL 32816-1364  
 tel. 407-823-5080  
[ramm@pegasus.cc.ucf.edu](mailto:ramm@pegasus.cc.ucf.edu)  
 Real and Complex analysis, Approximation Th.,  
 Fourier Analysis, Fuzzy Sets and Systems

23) Rainer Nagel  
 Arbeitsbereich Funktionalanalysis  
 Mathematisches Institut  
 Auf der Morgenstelle 10  
 D-72076 Tuebingen  
 Germany  
 tel. 49-7071-2973242  
 fax 49-7071-294322  
[rana@fa.uni-tuebingen.de](mailto:rana@fa.uni-tuebingen.de)  
 evolution equations, semigroups, spectral th.,  
 positivity

24) Panos M. Pardalos  
 Center for Appl. Optimization  
 University of Florida  
 303 Weil Hall  
 P.O. Box 116595  
 Gainesville, FL 32611-6595  
 tel. 352-392-9011  
[pardalos@ufl.edu](mailto:pardalos@ufl.edu)  
 Optimization, Operations Research



## BELGIUM

f.bastin@ulg.ac.be  
Functional Analysis, Wavelets

5) Yeol Je Cho  
Department of Mathematics Education  
College of Education  
Gyeongsang National University  
Chinju 660-701

## KOREA

tel. 055-751-5673 Office,  
055-755-3644 home,  
fax 055-751-6117  
yjcho@nongae.gsnu.ac.kr  
Nonlinear operator Th., Inequalities,  
Geometry of Banach Spaces

6) Sever S. Dragomir  
School of Communications and Informatics  
Victoria University of Technology  
PO Box 14428  
Melbourne City M.C  
Victoria 8001, Australia  
tel 61 3 9688 4437, fax 61 3 9688 4050  
sever.dragomir@vu.edu.au,  
sever@sci.vu.edu.au  
Math. Analysis, Inequalities, Approximation  
Th.,  
Numerical Analysis, Geometry of Banach  
Spaces,  
Information Th. and Coding

7) Angelo Favini  
Università di Bologna  
Dipartimento di Matematica  
Piazza di Porta San Donato 5  
40126 Bologna, ITALY  
tel. ++39 051 2094451  
fax. ++39 051 2094490  
favini@dm.unibo.it  
Partial Differential Equations, Control  
Theory,  
Differential Equations in Banach Spaces

8) Claudio A. Fernandez  
Facultad de Matematicas  
Pontificia Universidad Católica de Chile  
Vicuna Mackenna 4860  
Santiago, Chile  
tel. ++56 2 354 5922  
fax. ++56 2 552 5916  
cfernand@mat.puc.cl  
Partial Differential Equations,  
Mathematical Physics,  
Scattering and Spectral Theory

25) Svetlozar T. Rachev  
Dept. of Statistics and Applied Probability  
Program

University of California, Santa Barbara  
CA 93106-3110, USA  
tel. 805-893-4869  
rachev@pstat.ucsb.edu

## AND

Chair of Econometrics and Statistics  
School of Economics and Business Engineering  
University of Karlsruhe  
Kollegium am Schloss, Bau II, 20.12, R210  
Postfach 6980, D-76128, Karlsruhe, Germany  
tel. 011-49-721-608-7535  
rachev@lsoe.uni-karlsruhe.de  
Mathematical and Empirical Finance,  
Applied Probability, Statistics and Econometrics

26) John Michael Rassias  
University of Athens  
Pedagogical Department  
Section of Mathematics and Informatics  
20, Hippocratous Str., Athens, 106 80, Greece

Address for Correspondence

4, Agamemnonos Str.  
Aghia Paraskevi, Athens, Attikis 15342 Greece  
jrassias@primedu.uoa.gr  
jrassias@tellas.gr  
Approximation Theory, Functional Equations,  
Inequalities, PDE

27) Paolo Emilio Ricci  
Universita' degli Studi di Roma "La Sapienza"  
Dipartimento di Matematica-Istituto  
"G. Castelnuovo"  
P.le A. Moro, 2-00185 Roma, ITALY  
tel. ++39 0649913201, fax ++39 0644701007  
riccip@uniroma1.it, Paoloemilio.Ricci@uniroma1.it  
Orthogonal Polynomials and Special functions,  
Numerical Analysis, Transforms, Operational  
Calculus,  
Differential and Difference equations

28) Cecil C. Rousseau  
Department of Mathematical Sciences  
The University of Memphis  
Memphis, TN 38152, USA  
tel. 901-678-2490, fax 901-678-2480  
ccrousse@memphis.edu  
Combinatorics, Graph Th.,  
Asymptotic Approximations,  
Applications to Physics

29) Tomasz Rychlik

- 9) A.M.Fink  
Department of Mathematics  
Iowa State University  
Ames, IA 50011-0001, USA  
tel. 515-294-8150  
fink@math.iastate.edu  
Inequalities, Ordinary Differential Equations
- 10) Sorin Gal  
Department of Mathematics  
University of Oradea  
Str. Armatei Romane 5  
3700 Oradea, Romania  
galso@uoradea.ro  
Approximation Th., Fuzzyness, Complex Analysis
- 11) Jerome A. Goldstein  
Department of Mathematical Sciences  
The University of Memphis,  
Memphis, TN 38152, USA  
tel. 901-678-2484  
jgoldste@memphis.edu  
Partial Differential Equations, Semigroups of Operators
- 12) Heiner H. Gonska  
Department of Mathematics  
University of Duisburg  
Duisburg, D-47048  
Germany  
tel. 0049-203-379-3542 office  
gonska@informatik.uni-duisburg.de  
Approximation Th., Computer Aided Geometric Design
- 13) Dmitry Khavinson  
Department of Mathematical Sciences  
University of Arkansas  
Fayetteville, AR 72701, USA  
tel. (479) 575-6331, fax (479) 575-8630  
dmitry@uark.edu  
Potential Th., Complex Analysis, Holomorphic PDE,  
Approximation Th., Function Th.
- 14) Virginia S. Kiryakova  
Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
virginia@diogenes.bg  
Special Functions, Integral Transforms, Fractional Calculus
- 15) Hans-Bernd Knoop  
Institute of Mathematics  
Polish Academy of Sciences  
Chopina 12, 87100 Torun, Poland  
T.Rychlik@impan.gov.pl  
Mathematical Statistics, Probabilistic Inequalities
- 30) Bl. Sendov  
Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
bsendov@bas.bg  
Approximation Th., Geometry of Polynomials, Image Compression
- 31) Igor Shevchuk  
Faculty of Mathematics and Mechanics  
National Taras Shevchenko  
University of Kyiv  
252017 Kyiv  
UKRAINE  
shevchuk@univ.kiev.ua  
Approximation Theory
- 32) H.M. Srivastava  
Department of Mathematics and Statistics  
University of Victoria  
Victoria, British Columbia V8W 3P4  
Canada  
tel. 250-721-7455 office, 250-477-6960 home,  
fax 250-721-8962  
harimsri@math.uvic.ca  
Real and Complex Analysis, Fractional Calculus and Appl.,  
Integral Equations and Transforms, Higher Transcendental Functions and Appl., q-Series and q-Polynomials, Analytic Number Th.
- 33) Stevo Stevic  
Mathematical Institute of the Serbian Acad. of Science  
Knez Mihailova 35/I  
11000 Beograd, Serbia  
sstevic@ptt.yu; sstevo@matf.bg.ac.yu  
Complex Variables, Difference Equations, Approximation Th., Inequalities
- 34) Ferenc Szidarovszky  
Dept. Systems and Industrial Engineering  
The University of Arizona  
Engineering Building, 111  
PO. Box 210020  
Tucson, AZ 85721-0020, USA  
szidar@sie.arizona.edu  
Numerical Methods, Game Th., Dynamic Systems,

Institute of Mathematics  
Gerhard Mercator University  
D-47048 Duisburg  
Germany  
tel.0049-203-379-2676  
knoop@math.uni-duisburg.de  
Approximation Theory, Interpolation

16) Jerry Koliha  
Dept. of Mathematics & Statistics  
University of Melbourne  
VIC 3010, Melbourne  
Australia  
koliha@unimelb.edu.au  
Inequalities, Operator Theory,  
Matrix Analysis, Generalized Inverses

17) Mustafa Kulenovic  
Department of Mathematics  
University of Rhode Island  
Kingston, RI 02881, USA  
kulenm@math.uri.edu  
Differential and Difference Equations

18) Gerassimos Ladas  
Department of Mathematics  
University of Rhode Island  
Kingston, RI 02881, USA  
gladas@math.uri.edu  
Differential and Difference Equations

19) V. Lakshmikantham  
Department of Mathematical Sciences  
Florida Institute of Technology  
Melbourne, FL 32901  
e-mail: lakshmik@fit.edu  
Ordinary and Partial Differential  
Equations,  
Hybrid Systems, Nonlinear Analysis

20) Rupert Lasser  
Institut für Biomathematik & Biomertie, GSF  
-National Research Center for environment  
and health  
Ingolstaedter landstr.1  
D-85764 Neuherberg, Germany  
lasser@gsf.de  
Orthogonal Polynomials, Fourier Analysis,  
Mathematical Biology

Multicriteria Decision making,  
Conflict Resolution, Applications  
in Economics and Natural Resources  
Management

35) Gancho Tachev  
Dept. of Mathematics  
Univ. of Architecture, Civil Eng. and Geodesy  
1 Hr. Smirnenski blvd  
BG-1421 Sofia, Bulgaria  
gtt\_fte@uacg.bg  
Approximation Theory

36) Manfred Tasche  
Department of Mathematics  
University of Rostock  
D-18051 Rostock  
Germany  
manfred.tasche@mathematik.uni-rostock.de  
Approximation Th., Wavelet, Fourier Analysis,  
Numerical Methods, Signal Processing,  
Image Processing, Harmonic Analysis

37) Chris P. Tsokos  
Department of Mathematics  
University of South Florida  
4202 E. Fowler Ave., PHY 114  
Tampa, FL 33620-5700, USA  
profcp@math.usf.edu, profcp@chumal.cas.usf.edu  
Stochastic Systems, Biomathematics,  
Environmental Systems, Reliability Th.

38) Lutz Volkmann  
Lehrstuhl II für Mathematik  
RWTH-Aachen  
Templergraben 55  
D-52062 Aachen  
Germany  
volkm@math2.rwth-aachen.de  
Complex Analysis, Combinatorics, Graph Theory

### **EDITOR'S NOTE**

This special issue on “Applied Mathematics and Approximation Theory” contains expanded versions of articles that were presented in the international conference “Applied Mathematics and Approximation Theory 2008” ( AMAT 08), during October 11-13, 2008 at the University of Memphis, Memphis, Tennessee, USA.

All articles were refereed.

The organizer and Editor

George Anastassiou

# ITERATIVE RECONSTRUCTION AND STABILITY BOUNDS FOR SAMPLING MODELS

ERNESTO ACOSTA-REYES

**ABSTRACT.** This paper studies the reconstruction of a function  $f$  belonging to a shift-invariant space from the set of its non-uniformly distributed local sampled values. Here it is shown that if the set of sampling  $X = \{x_j\}_{j \in J}$  satisfies a necessary density conditions, then we can recover the function  $f$  from the set of its samples geometrically fast using an iterative algorithm. In addition, the algorithm is analyzed when the data is perturbed by noise, and it is proved that a small perturbation on the set of samples causes only a small change of the original function. Moreover, it is given an upper estimate of the rate of convergence of the algorithm. On the other hand, if we assume that  $X$  is a separated set, then it is shown that  $X$  is a set of sampling and explicit stability bounds are given.

## 1. INTRODUCTION

It is well known that in Sampling Theory there are two main goals (for an overview see [2], [4], [7]-[8], and [11]): First, given a class of functions on  $\mathbb{R}^d$ , to find conditions on the sampling set  $X = \{x_j\}_{j \in J}$ , where  $J$  is a countable index set, under which a function belonging to that class can be reconstructed uniquely and stably from its samples  $\{f(x_j)\}_{j \in J}$ . Second, to find efficient and fast numerical algorithms for recovering the function from its samples on  $X$ .

It is unrealistic to assume that the samples  $\{f(x_j)\}_{j \in J}$  can be measured exactly. For working with a more realistic model, we consider that our function(signal) belongs to a shift-invariant space  $V^p(\Phi)$ , for some  $1 \leq p \leq \infty$ , of the form

$$(1.1) \quad V^p(\Phi) = \left\{ \sum_{k \in \mathbb{Z}^d} C_k^T \Phi_k : C \in (\ell^p(\mathbb{Z}^d))^{(r)} \right\},$$

and that the samples of the signal have the form

$$g_{x_j}(f) = \int_{\mathbb{R}^d} f(x) d\mu_{x_j}(x),$$

---

*Key words and phrases.* Irregular sampling, Non-uniform sampling, Reconstruction, Fast algorithm, Shift-invariant spaces, Stability bounds.

where  $\mu = \{\mu_{x_j}\}_{j \in J}$  is a collection of finite complex Borel measures on  $\mathbb{R}^d$  that acts on the signal  $f$  in a neighborhood of  $x_j$  to produce the data  $\{g_{x_j}(f)\}_{j \in J}$ . The form of our sampled data generalizes the model presented by A. Aldroubi in [2] when for each  $j \in J$  the Radon-Nikodym derivative of  $\mu_{x_j}$  with respect to the Lebesgue measure on  $\mathbb{R}^d$  belongs to  $L^2(\mathbb{R}^d)$ . On the other hand, if the collection  $\mu$  consists of Dirac measures on  $\mathbb{R}^d$  concentrated at each point of  $X$ , then we obtain the model presented by A. Aldroubi and K. Gröchenig in [4].

In this paper we apply an iterative algorithm for recovering the signal  $f$  from its samples values  $\{g_{x_j}(f)\}_{j \in J}$  which uses the density properties of the set  $X$ , the support size conditions of the collection  $\mu$ , and the properties of the generator  $\Phi$  for  $V^p(\Phi)$ . Here we show that the sequence of functions generated using the algorithm converges to  $f$  geometrically fast. In [12], [14]-[18], this method was used for iterative reconstruction of band-limited signals, in [2] and [4], it was used for reconstructing functions belonging to shift-invariant spaces, and in [17] it was used for reconstructing signals belonging to a weighted multiply generated shift-invariant spaces. On the other hand, if  $X$  is assumed to be a separated set, then we show that  $X$  is also a set of sampling for  $V^p(\Phi)$  and  $\mu$ , and we give explicit stability bounds in terms of the rate of convergence of the algorithm, the generator for  $V^p(\Phi)$ , the bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$ , and the uniform upper bound for the total variations of the collection  $\mu$ . Moreover, it is given an upper estimate for the rate of convergence of the iterative algorithm.

The stability of the sampling reconstruction is analyzed when our local sampled data is perturbed by noise, and we show that a small perturbation of the sampled data  $\{g_{x_j}(f)\}_{j \in J}$  in the  $\ell^p(J)$  norm produces a small perturbation of our original function.

The remainder of this paper has been organized as follows. Section 2 introduces our sampling model, the definitions and notations that we shall work in this paper. The main results are presented in section 3, and we provide the proof of some of the results in section 4.

## 2. NOTATION AND PRELIMINARIES

In this section is introduced the sampling model we use in this paper, and the notations that will be used later.

The functions we are dealing with in this paper are functions  $f \in L^p(\mathbb{R}^d)$ , for some  $p \in [1, \infty]$  and  $d \in \mathbb{N}$ , which belong to a shift invariant space defined in (1.1), where  $\Phi = (\phi^1, \dots, \phi^r)^T$  is a vector of functions,  $\Phi_k = \Phi(\cdot - k)$ , and  $C = (c^1, \dots, c^r)^T$  is a vector of sequences belonging

## ITERATIVE RECONSTRUCTION AND STABILITY BOUNDS...

to  $(\ell^p(\mathbb{Z}^d))^{(r)}$ . Among the equivalent norms in  $(\ell^p(\mathbb{Z}^d))^{(r)}$  we choose

$$\|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}} = \sum_{i=1}^r \|c^i\|_{\ell^p(\mathbb{Z}^d)}.$$

Here it is assumed that the set  $\{\phi^1(\cdot - k), \dots, \phi^r(\cdot - k); k \in \mathbb{Z}^d\}$  generates an unconditional basis for  $V^p(\Phi)$ . In particular, we require that there exist constants  $0 < m_p \leq M_p < \infty$ , such that

$$(2.1) \quad m_p \|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}} \leq \left\| \sum_{k \in \mathbb{Z}^d} C_k^T \Phi_k \right\|_{L^p} \leq M_p \|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}}, \quad \forall C \in (\ell^p(\mathbb{Z}^d))^{(r)}.$$

The unconditional basis assumption (2.1) implies (see Theorem 2.4 in [4]) that the space  $V^p(\Phi)$  is a closed subspace of  $L^p(\mathbb{R}^d)$ .

Since we are interested in sampling in  $V^p(\Phi)$  we add an assumption that would make all the functions in these spaces continuous and, therefore, pointwise evaluations will be meaningful. Hence, we assume that the generator  $\Phi$  belongs to a Wiener-amalgam space  $(W_0^1)^{(r)}$  as defined below. For  $1 \leq p < \infty$ , a measurable function  $f$  belongs to  $W^p$  if it satisfies

$$(2.2) \quad \|f\|_{W^p} = \left( \sum_{k \in \mathbb{Z}^d} \operatorname{esssup}_{x \in [0,1]^d} |f(x+k)|^p \right)^{1/p} < \infty.$$

If  $p = \infty$ , a measurable function  $f$  belongs to  $W^\infty$  if it satisfies

$$(2.3) \quad \|f\|_{W^\infty} = \sup_{k \in \mathbb{Z}^d} \left\{ \operatorname{esssup}_{x \in [0,1]^d} |f(x+k)| \right\} < \infty.$$

Hence,  $W^\infty$  coincides with  $L^\infty(\mathbb{R}^d)$ . It is well known that for  $p \in [1, \infty]$ ,  $W^p$  is a Banach space (see [9]-[10]), and clearly  $W^p \subseteq L^p$ . By  $(W^p)^{(r)}$  we denote the space of vectors  $\Psi = (\psi^1, \dots, \psi^r)^T$  of  $W^p$ -functions with the norm

$$\|\Psi\|_{(W^p)^{(r)}} = \sum_{i=1}^r \|\psi^i\|_{W^p}.$$

The closed subspace of (vectors of) continuous functions in  $W^p$  (respectively,  $(W^p)^{(r)}$ ) will be denoted by  $W_0^p$  (or  $(W_0^p)^{(r)}$ ).

In this paper we are interested in average sampling performed by a countable collection of measures. We denote by  $\mathcal{M}(\mathbb{R}^d)$  the Banach space of finite complex Borel measures on  $\mathbb{R}^d$ . The norm on  $\mathcal{M}(\mathbb{R}^d)$  is given by  $\|\mu\| = \int_{\mathbb{R}^d} d|\mu|(y)$ , i.e., the total variation of a measure  $\mu$ .

Let  $J$  be a countable index set and  $X = \{x_j : j \in J\}$  be a subset of  $\mathbb{R}^d$ . The reconstruction problem in our sampling model consists of

finding the function  $f \in V^p(\Phi)$  from the knowledge of its samples

$$\left\{ g_{x_j}(f) = \int_{\mathbb{R}^d} f(x) d\mu_{x_j}(x) \right\}_{j \in J},$$

where  $\mu = \{\mu_{x_j}\}_{j \in J}$  is a countable collection of finite complex Borel measures on  $\mathbb{R}^d$  satisfying the following properties:

- (1) There exists  $a > 0$  such that  $\text{supp } \mu_{x_j} \subset x_j + [-a, a]^d$ , for all  $j \in J$ ,
- (2) There exists  $M > 0$  such that  $\|\mu_{x_j}\| \leq M$ , for all  $x_j \in X$ ; and
- (3)  $\int_{\mathbb{R}^d} d\mu_{x_j} = 1$ , for all  $j \in J$ .

**Definition 2.1.** Let  $1 \leq p \leq \infty$  and  $X = \{x_j : j \in J\}$  be a countable subset of  $\mathbb{R}^d$ . We say that  $X$  is a *set of sampling* for  $V^p(\Phi)$  and  $\mu = \{\mu_{x_j}\}_{j \in J}$  if there exist constants  $0 < A_p \leq B_p < \infty$  such that

$$(2.4) \quad A_p \|f\|_{L^p} \leq \|\{g_{x_j}(f)\}\|_{\ell^p(J)} \leq B_p \|f\|_{L^p}, \text{ for all } f \in V^p(\Phi).$$

$A_p$  and  $B_p$  are called the *stability bounds*.

*Remark 2.1.* If in the above definition we let  $p = 2$ , then applying Riesz representation theorem, it follows that (2.4) is the definition of frame. Thus,  $f$  can be reconstructed from its samples via dual frame expansion.

**Definition 2.2.** We say that  $X = \{x_j\}_{j \in J} \subset \mathbb{R}^d$  is *separated* if there exists  $\delta > 0$  such that  $\inf_{i, j \in J, i \neq j} |x_i - x_j| \geq \delta$ . The number  $\delta$  is called the *separation constant* of the set  $X$ .

**Definition 2.3.** A set  $X = \{x_j : j \in J\} \subset \mathbb{R}^d$  is  $\gamma$ -dense in  $\mathbb{R}^d$  if

$$\mathbb{R}^d = \bigcup_j B_r(x_j), \quad \forall r \geq \gamma,$$

where  $B_r(x_j) = \prod_{l=1}^d [x_j^l - r, x_j^l + r]$ .

**Definition 2.4.** A bounded partition of unit adapted to  $\{B_\gamma(x_j)\}_{j \in J}$  and associated with the sampling set  $X$  is a set of functions  $\{\beta_j\}_{j \in J}$  that satisfies:

- (1)  $0 \leq \beta_j \leq 1$ ,  $\forall j \in J$ ;
- (2)  $\text{supp } \beta_j \subset B_\gamma(x_j)$ ; and
- (3)  $\sum_{j \in J} \beta_j = 1$ .

Given a bounded partition of unit  $\{\beta_j\}_{j \in J}$  associated with the sampling set  $X$ , we define the operator  $A_X$  on  $V^p(\Phi)$  as follows

$$(2.5) \quad A_X f = \sum_{j \in J} g_{x_j}(f) \beta_j.$$



## ITERATIVE RECONSTRUCTION AND STABILITY BOUNDS...

The quasi-interpolant operator  $Q_X$  is defined on sequences  $c = \{c_j\}_{j \in J} \in \ell^p(J)$  by

$$(2.6) \quad Q_X c = \sum_{j \in J} c_j \beta_j.$$

If  $f \in W_0^p$ , we write

$$(2.7) \quad Q_X f = \sum_{j \in J} f(x_j) \beta_j$$

for the quasi-interpolant of the sequence  $c_j = f(x_j)$ .

*Remark 2.2.* Note that if  $\mu_{x_j} = \delta_{x_j}$ , for all  $j \in J$ , where  $\delta_{x_j}$  is the Dirac measure on  $\mathbb{R}^d$  concentrated at  $x_j$ , then  $A_X = Q_X$ .

## 3. MAIN RESULTS

In this section we collect the main results of our paper.

**Theorem 3.1.** *Let  $\Phi \in (W_0^1)^{(r)}$ ,  $1 \leq p \leq \infty$ , and  $P$  be a bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$ . Then there exists a density  $\gamma_0 = \gamma_0(\Phi, P, p) > 0$ , and  $a_0 = a_0(\Phi, P, p) > 0$  such that every  $f \in V^p(\Phi)$  can be recovered from the data  $\{g_{x_j}(f)\}_{j \in J}$  on any  $\gamma$ -dense set  $X = \{x_j\}_{j \in J}$  ( $0 < \gamma \leq \gamma_0$ ) for any support size condition (for  $\mu$ )  $0 < a \leq a_0$  by the following iterative algorithm:*

$$(3.1) \quad f_1 = P A_X f, \quad f_{n+1} = P A_X (f_n - f) + f_n.$$

*In this case the sequence  $\{f_n\}_{n \geq 1}$  converges to  $f$  in the  $W^p$  norm, hence both in the  $L^p(\mathbb{R}^d)$ , and uniformly. The convergence is geometric, that is,*

$$\|f_n - f\|_{L^p(\mathbb{R}^d)} \leq \|f_n - f\|_{W^p} \leq c_p \alpha^n \|f\|_{W^p},$$

*for some  $\alpha = \alpha(P, \gamma, a, \Phi, p) < 1$ , and for some  $0 < c_p < \infty$  independent of  $f$  and  $n \in \mathbb{N}$ .*

*Remark 3.1.* Notice that since  $\Phi \in (W_0^1)^{(r)}$ , then by Theorem 6.2 in [4] the existence of a bounded projection  $P$  is guaranteed for all  $p \in [1, \infty]$ , and in this case it is given by  $P f = \sum_{k \in \mathbb{Z}^d} \langle f, \tilde{\Phi}(\cdot - k) \rangle \Phi(\cdot - k)$ , where  $\{\tilde{\Phi}_k\}_{k \in \mathbb{Z}^d}$  is the canonical dual Riesz basis associated to  $\{\Phi_k\}_{k \in \mathbb{Z}^d}$ . Here  $\langle f, \tilde{\Phi} \rangle = (\langle f, \tilde{\phi}^1 \rangle, \dots, \langle f, \tilde{\phi}^r \rangle) \in \mathbb{C}^r$ ,  $\langle f, \tilde{\phi}^i \rangle = \int_{\mathbb{R}^d} f(z) \tilde{\phi}^i(z) dz$ , for  $1 \leq i \leq r$ , and  $\bar{z}$  denotes the complex conjugate of  $z$ .

*Remark 3.2.* Note that Theorem 4.1 in [2] is a Corollary of Theorem 3.1 when we let  $p = 2$  and  $r = 1$ .

Next result shows that if the hypothesis of Theorem 3.1 takes place, and  $X$  is also a separated set, then we obtain that  $X$  is a set of sampling for  $V^p(\Phi)$  and  $\mu$ , and explicit stability bounds are given.

**Theorem 3.2.** *Let  $\Phi \in (W_0^1)^{(r)}$  be given. Assume that  $X$  is separated with separation constant  $\delta > 0$ , and  $P$  is a bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$ . Then the following hold:*

- (1) *Given  $f \in V^p(\Phi)$ , the sequence  $\{f_n\}_{n \geq 1}$  defined by the algorithm (3.1) satisfies*

$$(3.2) \quad \|f_n\|_{L^p(\mathbb{R}^d)} \leq \left( \frac{1 + \alpha - \alpha^2}{1 - \alpha} \right) \|f\|_{L^p(\mathbb{R}^d)}, \quad \forall n \geq 1,$$

*where  $\alpha$  is rate of convergence of the algorithm (3.1).*

- (2)  *$X$  is a set of sampling for  $V^p(\Phi)$  and  $\mu$  with stability bounds given by*

$$(3.3) \quad A_p = \frac{1 - \alpha}{3^d \|P\|_{op} \mathcal{N}^{1/p'}},$$

*and*

$$(3.4) \quad B_p = \frac{M \mathcal{N}^{1/p} 3^{d/p} \|\Phi\|_{(W^1)^{(r)}}}{m_p},$$

*where  $\mathcal{N} = \mathcal{N}(\delta, p, d) = ([\frac{\sqrt{d}}{\delta}] + 1)^d$ ,  $\frac{1}{p} + \frac{1}{p'} = 1$ ,  $[t]$  denotes the biggest integer lower than or equal to  $t$ ,  $m_p$  is the lower bound constant in condition (2.1),  $\|P\|_{op}$  is the operator norm of  $P$ , and  $M > 0$  is the uniform upper bound for the total variations of the elements in the collection  $\mu$ .*

As a consequence of Theorem 3.1 and its proof, we obtain the following result which allows to find an estimate for the values of  $\gamma$  and  $a$  needed for the reconstruction algorithm (3.1). Moreover, Theorem 3.3 provides an upper estimate of the rate of convergence of the algorithm.

**Theorem 3.3.** *Assume that  $\Phi \in (W_0^1)^{(r)}$  and  $|\nabla \Phi| \in (W_0^1)^{(r)}$ , where  $|\nabla \Phi| = (|\nabla \phi^1|, \dots, |\nabla \phi^r|)^T$ , and  $\nabla \phi^i$  is the gradient of  $\phi^i$  for  $1 \leq i \leq r$ . Let  $M > 0$  be such that  $\|\mu_{x_j}\| \leq M$ , for all  $j \in J$ , and  $P$  be a bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$ . Then we have the following upper estimate for the rate of convergence of the algorithm (3.1):*

$$\alpha \leq \frac{\|P\|_{op}}{m_p} (\gamma(1+2[\gamma])^d + M((1+2[\gamma])^d + 2)a((1+2[a])^d)) \| |\nabla \Phi| \|_{(W^1)^{(r)}},$$

*where  $m_p$  is the lower bound constant given in (2.1), and  $[t]$  denotes the smallest integer bigger than or equal to  $t$ .*

### 3.1. Reconstructing in presence of noise.

Now we investigate the algorithm (3.1) in the case of noisy samples  $\{f'_j\}_{j \in J} \in \ell^p(J)$ , but we do not assume that and  $\{f'_j\}_{j \in J}$  are samples of a function  $f \in V^p(\Phi)$ . Then given  $\{\beta_j\}_{j \in J}$ , a bounded partition of unit associated with  $X$ , we use the initialization:

$$(3.5) \quad f_1 = P Q_X \{f'_j\}, \quad f_{n+1} = f_1 + (I - P A_X) f_n, \quad \forall n \geq 1,$$

and we have the following result.

**Theorem 3.4.** *Let  $\Phi \in (W_0^1)^{(r)}$ ,  $\{f'_j\}_{j \in J} \in \ell^p(J)$ , and  $P$  a bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$  be given. Then the algorithm (3.5) converges to a function  $f_\infty \in V^p(\Phi)$ , which satisfies  $P A_X f_\infty = P Q_X \{f'_j\}$ .*

As a consequence of Theorems 3.1 and 3.4, the next result shows the stability of the sampling-reconstruction.

**Theorem 3.5.** *Let  $\Phi \in (W_0^1)^{(r)}$ ,  $P$  a bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$  be given, and assume the  $X$  is a separated set. Let  $\{f'_j\}_{j \in J} \in \ell^p(J)$ , and  $f \in V^p(\Phi)$  with sampled values  $\{g_{x_j}(f)\}_{j \in J}$  be given. Then the following holds:*

$$(3.6) \quad \|f - f_\infty\|_{L^p} \leq \frac{3^d \mathcal{N}^{1/p'} \|P\|_{op}}{1 - \alpha} \|\{g_{x_j}(f) - f'_j\}\|_{\ell^p(J)},$$

where  $\mathcal{N} = ([\frac{\sqrt{d}}{\delta}] + 1)^d$ ,  $\frac{1}{p} + \frac{1}{p'} = 1$ ,  $\alpha = \|I - P A_X\|_{op}$ ,  $f_\infty \in V^p(\Phi)$  is the function given in Theorem 3.4, and  $\delta > 0$  is the separation constant of the set  $X$ .

## 4. PROOFS

### 4.1. Auxiliary results.

We begin this section with three results that are needed for the main proofs.

The next Lemma collects basic facts about Wiener amalgam spaces, and shift-invariant spaces. For a proof of this Lemma see Proposition 4.2 in [1].

**Lemma 4.1.** *Let  $\Phi \in (W^1)^{(r)}$ ,  $f = \sum_{k \in \mathbb{Z}^d} C_k^T \Phi_k$ , where  $C \in (\ell^p(\mathbb{Z}^d))^{(r)}$ , and  $\Phi_k = \Phi(\cdot - k)$ , for all  $k \in \mathbb{Z}^d$ . Then the following hold:*

$$(4.1) \quad V^p(\Phi) \subset W_0^p, \text{ for all } 1 \leq p \leq \infty,$$

if  $\Phi \in (W_0^1)^{(r)}$ .

$$(4.2) \quad \|f\|_{W^p} \leq \|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}} \|\Phi\|_{(W^1)^{(r)}}.$$

We also need the following Lemma which will be stated without proof (see Lemmas 5.1 and 5.2 in [2], and Lemma 8.1 in [4] and the references therein).

**Lemma 4.2.** *Let  $\Phi \in (W_0^1)^{(r)}$ , and  $f = \sum_{k \in \mathbb{Z}^d} C_k^T \Phi_k$ , where  $C \in (\ell^p(\mathbb{Z}^d))^{(r)}$ . Then:*

- (1) *The oscillation  $\text{osc}_\gamma(f)$  belongs to  $W^p$ .*
- (2) *The oscillation  $\text{osc}_\gamma \Phi$  satisfies*

$$(4.3) \quad \|\text{osc}_\gamma \Phi\|_{(W^1)^{(r)}} \leq ((1 + 2\lceil \gamma \rceil)^d + 1) \|\Phi\|_{(W^1)^{(r)}},$$

*and  $\|\text{osc}_\gamma \Phi\|_{(W^1)^{(r)}} \rightarrow 0$  as  $\gamma \rightarrow 0$ .*

- (3) *If  $|\nabla \Phi| \in (W_0^1)^{(r)}$ , then*

$$(4.4) \quad \|\Phi\|_{(W^1)^{(r)}} \leq \gamma(2\lceil \gamma \rceil + 1)^d \|\nabla \Phi\|_{(W^1)^{(r)}}$$

- (4) *The oscillation  $\text{osc}_\gamma(f)$  satisfies*

$$(4.5) \quad \|\text{osc}_\gamma(f)\|_{W^p} \leq \|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}} \|\text{osc}_\gamma \Phi\|_{(W^1)^{(r)}}, \quad \forall C \in (\ell^p(\mathbb{Z}^d))^{(r)}.$$

*In particular,  $\|\text{osc}_\gamma(f)\|_{W^p} \rightarrow 0$  as  $\gamma \rightarrow 0$ . Moreover,*

$$(4.6) \quad \|\mathbf{Q}_X f\|_{L^p} \leq \|\mathbf{Q}_X f\|_{W^p} \leq ((1 + 2\lceil \gamma \rceil)^d + 2) \|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}} \|\Phi\|_{(W^1)^{(r)}}, \quad \forall C \in (\ell^p(\mathbb{Z}^d))^{(r)}.$$

**Lemma 4.3.** *Let  $\Phi \in (W_0^1)^{(r)}$  be given. Let  $P$  be a bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$ . Then there exist  $\gamma_0 = \gamma_0(\Phi, P, p) > 0$ , and  $a_0 = a_0(\Phi, P, p) > 0$  such that for any  $0 < a \leq a_0$ , the operator  $I - P A_X$  is a contraction on  $V^p(\Phi)$  for any  $\gamma$ -dense set  $X$  with  $0 < \gamma \leq \gamma_0$ .*

*Proof.* Let  $P$  be a bounded projection from  $L^p(\mathbb{R}^d)$  onto  $V^p(\Phi)$ , and  $f = \sum_{k \in \mathbb{Z}^d} C_k^T \Phi_k$ , where  $C \in (\ell^p(\mathbb{Z}^d))^{(r)}$  be given. Then

$$\begin{aligned} |f(x) - (\mathbf{Q}_X f)(x)| &= \left| f(x) - \sum_{j \in J} f(x_j) \beta_j(x) \right| \\ &= \left| \sum_{j \in J} (f(x) - f(x_j)) \beta_j(x) \right| \\ &\leq \sum_{j \in J} |f(x) - f(x_j)| \beta_j(x) \\ &\leq \text{osc}_\gamma(f)(x) \sum_{j \in J} \beta_j(x) = \text{osc}_\gamma(f)(x). \end{aligned}$$

## ITERATIVE RECONSTRUCTION AND STABILITY BOUNDS...

From this pointwise estimate and (4.5) we obtain

$$\begin{aligned} \|f - Q_X f\|_{L^p} &\leq \|f - Q_X f\|_{W^p} \\ &\leq \|\operatorname{osc}_\gamma(f)\|_{W^p} \leq \|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}} \|\operatorname{osc}_\gamma \Phi\|_{(W^1)^{(r)}} \\ &\leq \frac{1}{m_p} \|f\|_{L^p} \|\operatorname{osc}_\gamma \Phi\|_{(W^1)^{(r)}}, \end{aligned}$$

where we have used condition (2.1) in the last inequality. Consequently,

$$(4.7) \quad \|f - Q_X f\|_{L^p} \leq \frac{1}{m_p} \|f\|_{L^p} \|\operatorname{osc}_\gamma \Phi\|_{(W^1)^{(r)}}.$$

On the other hand,

$$\begin{aligned} |(Q_X f - A_X f)(x)| &= \left| \sum_{j \in J} (f(x_j) - g_{x_j}(f)) \beta_j(x) \right| \\ &= \left| \sum_{j \in J} \left( \int_{\mathbb{R}^d} (f(x_j) - f(z)) d\mu_{x_j}(z) \right) \beta_j(x) \right| \\ &\leq \sum_{j \in J} \int_{\mathbb{R}^d} |f(x_j) - f(z)| d|\mu_{x_j}|(z) \beta_j(x) \\ &\leq \sum_{j \in J} \operatorname{osc}_a(f)(x_j) \beta_j(x) \int_{\mathbb{R}^d} d|\mu_{x_j}|(z) \\ &\leq M \sum_{j \in J} \operatorname{osc}_a(f)(x_j) \beta_j(x) \\ &\leq M \sum_{j \in J} \left( \sum_{i=1}^r \sum_{k \in \mathbb{Z}^d} |c_k^i| \operatorname{osc}_a(\phi^i)(x_j - k) \right) \beta_j(x). \end{aligned}$$

By using Lemma 4.2, condition (2.1), Triangular inequality, and the above pointwise estimate, we have

$$(4.8) \quad \|Q_X f - A_X f\|_{L^p} \leq \frac{M}{m_p} ((1 + 2\lceil \gamma \rceil)^d + 2) \|\operatorname{osc}_a \Phi\|_{(W^1)^{(r)}} \|f\|_{L^p}.$$

Since  $f \in V^p(\Phi)$ , then  $P f = f$ . Therefore,

$$\begin{aligned} \|f - P A_X f\|_{L^p} &\leq \|P f - P Q_X f\|_{L^p} + \|P Q_X f - P A_X f\|_{L^p} \\ &\leq \|P\|_{op} \|f - Q_X f\|_{L^p} + \|P\|_{op} \|Q_X f - A_X f\|_{L^p}. \end{aligned}$$

Using now (4.7), (4.8), and the above inequality we get

$$(4.9) \quad \|f - P A_X f\|_{L^p} \leq \frac{\|P\|_{op}}{m_p} \left( \|\operatorname{osc}_\gamma \Phi\|_{(W^1)^{(r)}} + M((1 + 2\lceil \gamma \rceil)^d + 2) \|\operatorname{osc}_a \Phi\|_{(W^1)^{(r)}} \right) \|f\|_{L^p}.$$

Let  $0 < \epsilon < \frac{m_p}{\|P\|_{op}}$  be given. Since  $\|\text{osc}_\gamma \Phi\|_{(W^1)^{(r)}} \rightarrow 0$  as  $\gamma \rightarrow 0^+$ , then there exists  $\gamma_0 = \gamma_0(\epsilon, \Phi, P, p) > 0$ , and  $a_0 = a_0(\epsilon, \Phi, P, p) > 0$  such that

$$\|\text{osc}_\gamma \Phi\|_{(W^1)^{(r)}} \leq \frac{\epsilon}{2}, \quad \text{for all } 0 < \gamma \leq \gamma_0,$$

and

$$M((1 + 2\lceil \gamma \rceil)^d + 2)\|\text{osc}_a \Phi\|_{(W^1)^{(r)}} \leq \frac{\epsilon}{2}, \quad \text{for all } 0 < a \leq a_0.$$

If we choose  $\gamma_0$  and  $a_0$  so that for any  $0 < \gamma \leq \gamma_0$ , and  $0 < a \leq a_0$  we have

$$\|f - P A_X f\|_{L^p} \leq \frac{\epsilon \|P\|_{op}}{m_p} \|f\|_{L^p},$$

then the conclusion of the Lemma follows.  $\square$

#### 4.2. Proofs for Section 3.

Now we are ready to prove our main results.

##### Proof of Theorem 3.1.

*Proof.* Let  $e_n = f - f_n$  be the error after  $n$  iterations of the algorithm (3.1). Then the sequence  $\{e_n\}_{n \in \mathbb{N}}$  satisfies

$$\begin{aligned} e_{n+1} &= f - f_{n+1} = f - f_n - P A_X (f - f_n) \\ &= (I - P A_X)(f - f_n) = (I - P A_X)(e_n). \end{aligned}$$

By using Lemma 4.3, there exist a density  $\gamma_0 > 0$ , and  $a_0 > 0$  such that for any  $0 < \gamma \leq \gamma_0$ , and  $0 < a \leq a_0$ ,  $I - P A_X$  is a contraction on  $V^p(\Phi)$ . Therefore, by taking  $\alpha := \|I - P A_X\|_{op} < 1$ , we have

$$\|e_{n+1}\|_{L^p} \leq \alpha \|e_n\|_{L^p},$$

and by induction it follows that

$$(4.10) \quad \|e_{n+1}\|_{L^p} \leq \alpha^{n+1} \|f\|_{L^p},$$

and  $\|e_n\|_{L^p} \rightarrow 0$  geometrically fast. Since for  $V^p(\Phi)$  the  $W^p$  norm, and the  $L^p$  norm are equivalent, then (4.10) also holds in the  $W^p$  norm and uniformly on  $\mathbb{R}^d$ , and Theorem 3.1 is proved.  $\square$

##### Proof of Theorem 3.2.

*Proof.* Let us prove (3.2). Note that by hypothesis and Lemma 4.3 we have that there exists  $\gamma_0 > 0$  such that  $I - P A_X$  is a contraction for any  $\gamma$ -dense set  $X$  with  $0 < \gamma \leq \gamma_0$ . Hence,  $\alpha = \|I - P A_X\|_{op} < 1$ , and thus, the operator  $P A_X$  is invertible on  $V^p(\Phi)$ . It is not hard to show that  $P A_X$  and  $(P A_X)^{-1}$  satisfy:

$$(4.11) \quad 1 - \alpha \leq \|P A_X\|_{op} \leq 1 + \alpha,$$

and

$$(4.12) \quad \frac{1}{1+\alpha} \leq \|(P A_X)^{-1}\|_{op} \leq \frac{1}{1-\alpha}.$$

Let  $f \in V^p(\Phi)$  be given. Since  $\{f_n\}_{n \geq 1}$  given by algorithm (3.1) satisfies  $f_n = f_1 + e_1 + e_2 + \dots + e_{n-1}$  for  $n \geq 2$ ,  $f_1 = P A_X f$ , and since  $\{e_n\}_{n \geq 1}$  satisfies  $e_n = (I - P A_X)e_{n-1}$ , for  $n \geq 1$ , then we get

$$f_n = f_1 + \sum_{l=1}^n (I - P A_X)^l f.$$

Hence,

$$\begin{aligned} \|f_n\|_{L^p} &\leq \|f_1\|_{L^p} + \|f\|_{L^p} \sum_{l=1}^n \|I - P A_X\|_{op}^l \\ &\leq \left( \|P A_X\|_{op} + \sum_{l=1}^n \alpha^l \right) \|f\|_{L^p} \\ &\leq \left( 1 + \alpha + \frac{\alpha}{1-\alpha} \right) \|f\|_{L^p} = \left( \frac{1+\alpha-\alpha^2}{1-\alpha} \right) \|f\|_{L^p}, \end{aligned}$$

and we obtain (3.2). Let us show (3.3). Let  $f \in V^p(\Phi)$  be given. Then by Lemma 4.3, there exists  $\gamma_0 > 0$  such that the operator  $I - P A_X$  is a contraction on  $V^p(\Phi)$  for any  $\gamma$ -dense set  $X$  with  $0 < \gamma \leq \gamma_0$ . Hence,  $\alpha = \|I - P A_X\|_{op} < 1$ , the operator  $P A_X$  is invertible, and (4.12) takes place. On the other hand, from the definition of the operators  $A_X$  and  $Q_X$ , it follows that  $A_X f = Q_X \{g_{x_j}(f)\}$ , and thus,  $P A_X f = P Q_X \{g_{x_j}(f)\}$ . Therefore,  $f = (P A_X)^{-1} P Q_X \{g_{x_j}(f)\}$ . Consequently,

$$\begin{aligned} \|f\|_{L^p} &\leq \|(P A_X)^{-1}\|_{op} \|P\|_{op} \|Q_X\|_{op} \|\{g_{x_j}(f)\}\|_{\ell^p(J)} \\ &\leq \frac{\|P\|_{op} \|Q_X\|_{op}}{1-\alpha} \|\{g_{x_j}(f)\}\|_{\ell^p(J)}. \end{aligned}$$

In order to complete the proof of (3.3), we need an upper estimate for  $\|Q_X\|_{op}$ . Let  $\chi$  be the characteristic function of the set  $B_\gamma(0) + [0, 1]^d$ . Clearly, we may assume without loss of generality that  $0 < \gamma < 1$ . Since  $0 \leq \beta_j \leq 1$ , and  $\text{supp } \beta_j \subset B_\gamma(x_j)$ , then for all  $x_j \in k + [0, 1]^d$ ,  $\beta_j(x) \leq \chi(x - k)$ . Therefore,

$$|Q_X c| = \left| \sum_{j \in J} c_j \beta_j(x) \right| \leq \sum_{k \in \mathbb{Z}^d} \left( \sum_{j: x_j \in k + [0, 1]^d} |c_j| \right) \chi(x - k),$$

and by (4.2) in Lemma 4.1, we have

$$\|Q_X c\|_{W^p} \leq \left( \sum_{k \in \mathbb{Z}^d} \left( \sum_{j: x_j \in k + [0,1]^d} |c_j| \right)^p \right)^{1/p} \|\chi\|_{W^1}.$$

Since  $X$  is a separated set with separation constant  $\delta > 0$ , then there are at most  $\mathcal{N} = \mathcal{N}(\delta, p, d) = ([\frac{\sqrt{d}}{\delta}] + 1)^d$  sampling points  $x_j$  in each cube  $k + [0, 1]^d$ . By applying Hölder's inequality, we get

$$\left( \sum_{j: x_j \in k + [0,1]^d} |c_j| \right)^p \leq \mathcal{N}^{p/p'} \sum_{j: x_j \in k + [0,1]^d} |c_j|^p,$$

where  $\frac{1}{p} + \frac{1}{p'} = 1$ . Consequently,

$$\|Q_X c\|_{W^p} \leq \mathcal{N}^{1/p'} \|\{c_j\}\|_{\ell^p(J)} \|\chi\|_{W^1}.$$

We leave to the reader the proof of  $\|\chi\|_{W^1} = 3^d$ . Therefore,

$$\|Q_X c\|_{W^p} \leq 3^d \mathcal{N}^{1/p'} \|\{c_j\}\|_{\ell^p(J)},$$

and

$$(4.13) \quad \|Q_X\|_{op} \leq 3^d \mathcal{N}^{1/p'}.$$

Hence,

$$\frac{1 - \alpha}{3^d \|P\|_{op} \mathcal{N}^{1/p'}} \|f\|_{L^p} \leq \|\{g_{x_j}(f)\}\|_{\ell^p(J)}, \quad \text{for all } f \in V^p(\Phi).$$

Let us show (3.4). Note that

$$\begin{aligned} \sum_{x_j \in k + [0,1]^d} |g_{x_j}(f)|^p &= \sum_{x_j \in k + [0,1]^d} \left| \int_{\mathbb{R}^d} f(z) d\mu_{x_j}(z) \right|^p \\ &\leq \sum_{x_j \in k + [0,1]^d} \|\mu_{x_j}\|^p \left( \int_{\mathbb{R}^d} |f(z)| \frac{d|\mu_{x_j}|(z)}{\|\mu_{x_j}\|} \right)^p \\ &\leq \sum_{x_j \in k + [0,1]^d} \|\mu_{x_j}\|^p \int_{\mathbb{R}^d} |f(z)|^p \frac{d|\mu_{x_j}|(z)}{\|\mu_{x_j}\|} \\ &\leq M^p \sum_{x_j \in k + [0,1]^d} \text{esssup}_{z \in x_j + [-a,a]^d} |f(z)|^p. \end{aligned}$$

Since  $X$  is a separated set, then there exist at most  $\mathcal{N} = \mathcal{N}(\delta, p, d) = ([\frac{\sqrt{d}}{\delta}] + 1)^d$  sampling points in each cube  $k + [0, 1]^d$ . Assuming without loss of generality that  $0 < a \leq 1$ , then

$$\sum_{x_j \in k + [0,1]^d} |g_{x_j}(f)|^p \leq M^p 3^d \mathcal{N} \text{esssup}_{z \in k + [0,1]^d} |f(z)|^p.$$



## ITERATIVE RECONSTRUCTION AND STABILITY BOUNDS...

Consequently, by taking the sum over  $k \in \mathbb{Z}^d$  in the above inequality, we obtain:

$$\begin{aligned} \|\{g_{x_j}(f)\}\|_{\ell^p(J)} &\leq M\mathcal{N}^{1/p}3^{d/p}\|f\|_{W^p} \\ &\leq M\mathcal{N}^{1/p}3^{d/p}\|C\|_{(\ell^p(\mathbb{Z}^d))^{(r)}}\|\Phi\|_{(W^1)^{(r)}} \\ &\leq \frac{M\mathcal{N}^{1/p}3^{d/p}\|\Phi\|_{(W^1)^{(r)}}}{m_p}\|f\|_{L^p}, \end{aligned}$$

and Theorem 3.2 is proved.  $\square$

**Proof of Theorem 3.3.**

*Proof.* The proof of Theorem 3.3 is a straightforward consequence of (4.3), (4.4), and (4.9).  $\square$

**Proof of Theorem 3.4.**

*Proof.* Assume the hypothesis of Theorem 3.4 holds. From Lemma 4.3, the operator  $I - P A_X$  is a contraction. Consequently, the sequence  $\{f_n\}_{n \geq 1}$  defined by algorithm (3.5) converges to a function  $f_\infty \in V^p(\Phi)$ . By taking limits in both sides of (3.5) as  $n \rightarrow \infty$ , we have:

$$f_\infty = f_1 + (I - P A_X)f_\infty.$$

Therefore,  $f_1 - P A_X f_\infty = 0$ . Taking into account that  $f_1 = P Q_X \{f'_j\}$ , then the conclusion of Theorem 3.4 follows.  $\square$

**Proof of Theorem 3.5.**

*Proof.* Assume that the hypothesis of Theorem 3.5 holds. By Lemma 4.3, there exists  $\gamma_0 > 0$  such that the operator  $I - P A_X$  is a contraction on  $V^p(\Phi)$  for any  $\gamma$ -dense set  $X$  with  $0 < \gamma \leq \gamma_0$ . Hence,  $\alpha = \|I - P A_X\|_{op} < 1$ , the operator  $P A_X$  is invertible, and (4.12) takes place. On the other hand, from the definition of the operators  $A_X$  and  $Q_X$ , it follows that  $A_X f = Q_X \{g_{x_j}(f)\}$ , and thus,  $P A_X f = P Q_X \{g_{x_j}(f)\}$ . Therefore,  $f = (P A_X)^{-1} P Q_X \{g_{x_j}(f)\}$ . By applying now Theorem 3.4, then there exists a function  $f_\infty \in V^p(\Phi)$  such that  $P A_X f_\infty = P Q_X \{f'_j\}$ . Hence,  $f_\infty = (P A_X)^{-1} P Q_X \{f'_j\}$ . Consequently,

$$\begin{aligned} \|f - f_\infty\|_{L^p} &\leq \|(P A_X)^{-1}\|_{op} \|P Q_X\|_{op} \|\{g_{x_j}(f) - f'_j\}\|_{\ell^p(J)} \\ &\leq \frac{\|P Q_X\|_{op}}{1 - \alpha} \|\{g_{x_j}(f) - f'_j\}\|_{\ell^p(J)} \\ &\leq \frac{\|P\|_{op} \|Q_X\|_{op}}{1 - \alpha} \|\{g_{x_j}(f) - f'_j\}\|_{\ell^p(J)}. \end{aligned}$$

Now the conclusion of Theorem 3.5 follows by using (4.13).  $\square$

### Acknowledgments

I would like to thank Professor Akram Aldroubi for his valuable suggestions, insightful remarks, and editorial advice.

### REFERENCES

- [1] E. Acosta-Reyes, A. Aldroubi, and I. Krishtal, On stability of sampling-reconstruction models, to appear in *Adv. in Comp. Math.*, 2008.
- [2] A. Aldroubi, Non-uniform weighted average sampling and reconstruction in shift-invariant and wavelet spaces, *Appl. Comput. Harmon. Anal.* **13**, 2002, 151-161.
- [3] A. Aldroubi, K. Gröchenig, Beurling-Landau-type theorems for non-uniform sampling in shift invariant spline spaces, *J. Fourier Anal. Appl.* **6**, 2000, 93-103.
- [4] A. Aldroubi, K. Gröchenig, Nonuniform sampling and reconstruction in shift-invariant spaces, *SIAM Rev.* **43**, 2001, 585-620.
- [5] J. J. Benedetto, Irregular sampling and frames. In: *Wavelets: A Tutorial in Theory and Applications* (C.K. Chui, Ed.) Boston: Academic Press, 1992, 445-507.
- [6] P. L. Butzer, J. Lei, Approximation of signals using measured sampled values error analysis. *Comm. Appl. Anal.*, **4**, 2000, 245-255.
- [7] P. L. Butzer, W. Splettöber, R. L. Sten, The sampling theorem and linear prediction in signal analysis, *Jahresber. Deutsch. Math.-Verein* **90**, 1988, 1-70.
- [8] W. Chen, S. Itoh, and J. Shiki, On Sampling in Shift Invariant Spaces, *IEEE Trans. on Information Theory*, **48**, no. 10, 2002, 2802-2809.
- [9] H. G. Feichtinger, Generalized amalgams with applications to Fourier Transform., *Can. J. Math.* **42**(3), 1990, 395-409.
- [10] H. G. Feichtinger, Wiener amalgam over Euclidean spaces and some of their applications, in: K. Jarosz (Ed.) *Proc. Conf. Function Spaces*, Edwardsville, IL, USA, 1990, in: *Lecture Notes in Pure and Appl. Math.*, vol. **136**, 1992, 107-121.
- [11] H. G. Feichtinger, K. Gröchenig, Theory and practice of irregular sampling. In *wavelets: Mathematics and Applications* (J. Benedetto, M. Frazier, eds.) Boca Raton, FL: CRC Press, 305-363.
- [12] K. Gröchenig, Reconstruction algorithms in irregular sampling, *Math. Comput.*, **59**, 1992, 181-194.
- [13] F. Maruasti and M. Analoui, Recovery signals from nonuniform sampling using iterative methods, *Proc. Internat. Sympos. Circuits Systems*, Portland, OR, 1987.
- [14] K. D. Sauer and J. P. Allebach, Iterative reconstruction of band-limited images from nonuniformly spaced samples, *IEEE Trans. Circuits and Systems*, **34** 1987, 1497-1506.
- [15] W. Sun, X. Zhou, Reconstruction of bandlimited functions from local averages, *Const. Approx.* **18**, 2002, 205-222.
- [16] R. G. Wiley, Recovery of band-limited signals from unequally spaced samples., *IEEE Trans. Comm.*, **26**, 1978, 135-137.

## ITERATIVE RECONSTRUCTION AND STABILITY BOUNDS...

- [17] J. Xian and S. Li, Sampling set conditions in weighted multiply generated shift-invariant spaces and their applications, *Appl. Comput. Harmon. Anal.* **23**, 2007, 171-180.
- [18] S. Yeh and H. Stark, Iterative and one-step reconstruction from nonuniform samples by convex projections, *J. Opt. Soc. Amer.* **A7**, 1990, 491-499.
- [19] D. C. Youla and H. Webb, Image restoration by method of convex projections: Part 1-Theory, *IEEE Trans. Med. Image* **1**, 1982, 81-94.

DEPT. OF MATHEMATICS, VANDERBILT UNIVERSITY, NASHVILLE, TN 37240,  
USA., EMAIL: ERNESTO.ACOSTA@VANDERBILT.EDU

# GLOBAL EXISTENCE AND BLOW UP FOR SOLUTIONS TO HIGHER ORDER BOUSSINESQ SYSTEMS

DÉ GODEFROY AKMEL

*Laboratoire de Mathématiques Appliquées, UFRMI, Université d'Abidjan Cocody, 22 BP 582  
Abidjan 22, Cote d'Ivoire. E-mail akmelde@yahoo.fr*

ABSTRACT. In this paper, we deal with higher order Boussinesq systems of equations in one dimension, that has been presented by Bona, Chen and Saut in [1]. We show that the solutions of these systems of equations with a nonlinear power  $\alpha \geq 1$  are global and decay in time for small initial data, and we show also that they blow-up in finite time.

*Keywords* decay in time, Boussinesq equation, blow-up

*AMS Subject Classification:* 35B40; 35Q10; 35Q20

## 1. INTRODUCTION

In this paper, consideration is given to the higher order Boussinesq system

$$(1.1) \quad \eta_t - b\eta_{xxt} + b_2\eta_{xxxxt} + u_x + au_{xxx} = -(f_1(\eta, u))_x + b_2(f_1(\eta, u))_{xxx},$$

$$(1.2) \quad u_t - du_{xxt} + d_2u_{xxxxt} + \eta_x + a\eta_{xxx} = -(f_2(\eta, u))_x + d_2(f_2(\eta, u))_{xxx},$$

with initial data  $\eta(x, 0) = \eta_0(x)$ ,  $u(x, 0) = u_0(x)$ , and where the nonlinearities  $f_1(\eta, u) = \eta^\alpha u + c_1\eta u^\alpha$ ,  $f_2(\eta, u) = \eta^{\alpha+1} + c_2u^{\alpha+1}$ , with  $\alpha \geq 1$  integer,  $c_1, c_2 \in \mathbb{R}$ ; the constants  $a, b, d, b_2, d_2 \in \mathbb{R}$  verify some conditions below, and  $x \in \mathbb{R}$ ,  $t \geq 0$ . The system (1.1)-(1.2) describes the propagation of surface water waves. Here, the independent variable,  $x$ , is proportional to distance in the directional of propagation while  $t$  is proportional to elapsed time. The quantity  $\eta(x, t) + h_0$  corresponds to the total depth of the liquid at the point  $x$  at time  $t$ , where  $h_0$  is the indisturb water depth. The variable  $u(x, t)$  represents the horizontal velocity at the point  $(x, t)$ . Our study here is devoted to the following particular cases for the system (1.1)-(1.2):

$C1 = [b > 0, d \geq 0, b_2 = 0, d_2 > 0, \text{ with } a > 0 : \text{case } C1-1 \text{ or } a = 0 : \text{case } C1-2]$

$C2 = [b \geq 0, d > 0, b_2 > 0, d_2 = 0, \text{ with } a > 0 : \text{case } C2-1 \text{ or } a = 0 : \text{case } C2-2]$

$C3 = [a = 0, b \geq 0, d \geq 0, b_2 > 0, d_2 > 0]$

$C4 = [b > 0, d > 0, b_2 = d_2 = 0, \text{ with } a > 0 : \text{case } C4-1 \text{ or } a = 0 : \text{case } C4-2]$

$$C5 = \begin{cases} C5-1 : a \neq 0, b > 0, d > 0, b_2 = 0, d_2 \geq 0, \text{ or} \\ C5-2 : a \neq 0, b > 0, d > 0, b_2 \geq 0, d_2 = 0. \end{cases}$$

A local existence result has been obtained by Bona *et al.* in [1] for the full I.V.P (1.1)-(1.2) when  $\alpha = 1$  and  $a < 0, b \geq 0, d \geq 0, b_2 > 0, d_2 > 0, c_1, c_2 \in \mathbb{R}$ . On the other hand, the case (C4-2) which corresponds to the purely BBM-Boussinesq and the case (C4-1) have been shown by the same authors to be locally well posed

in  $\mathbf{H}^s(\mathbf{R})$   $s \geq 0$  ( see Bona *et all.* [1]). Global existence in  $\mathbf{H}^s(\mathbf{R})$   $s \geq 1$  has been proved in [1] for (1.1)-(1.2) under the condition  $a < 0$ ,  $b = d > 0$ ,  $b_2 = d_2 = 0$  with  $\alpha = 1$ . Note that this condition is include in the case (C5). Furthermore, when studying the system (1.1)-(1.2) under the restriction (C4-1) with  $\alpha = 1$  and with complete or partial dissipation, Chen and Goubet showed in [2] that the solution of this system decays to zero when  $t$  goes to  $+\infty$ .

Our study here is concerned with the asymptotic behavior of the solution to the Boussinesq system (1.1)-(1.2). In the first part, we show that the solution of (1.1)-(1.2) under the restrictions (C1), (C2), (C3) or (C4), and with  $\alpha > 5$  (for the cases C1-1, C2-1, C4-1) or  $\alpha > 9$ , (for the cases C1-2, C2-2, C3, C4-2), decays to zero when  $t$  goes to  $+\infty$ .

In the second part, we show that the solution of (1.1)-(1.2) under the restriction (C5) with  $a = -b = -d_2$  and where  $\alpha \geq 1$ , blows-up in a finite time.

**1.1. Notation.** The notation  $\|\cdot\|_{r,p}$  is used to denote the norm in  $\mathbb{L}_r^p (= \mathbb{H}^{r,p})$  such that if we set  $J^r = (1 - \frac{\partial^2}{\partial x^2})^{r/2}$ , then for  $u \in \mathbb{L}_r^p(\mathbb{R})$ ,  $\|u\|_{r,p} = \|u\|_{\mathbb{L}_r^p} = \|J^r u\|_{\mathbb{L}^p} < \infty$ . Also,  $|\cdot|_p$  instead of  $\|\cdot\|_{0,p}$  denotes the norm in  $\mathbb{L}^p$ , and  $\mathbb{H}^s$  is used instead of  $\mathbb{L}_s^2$ . Throughout the paper,  $c$  represents a generic constant independent of  $t$  and  $x$ . The Fourier transform of a function  $f$  is denoted by  $\hat{f}(\xi)$  or  $\mathcal{F}(f)(\xi)$  and  $\mathcal{F}^{-1}(f) \equiv \check{f}$  denotes the inverse Fourier transform of  $f$ .

## 2. LOCAL EXISTENCE IN TIME.

In this section, we study the local existence in time for the solution to the Cauchy problem associated to (1.1)-(1.2) under the conditions C1, C2, C3, C4, or C5 above. We prove the following theorem:

**Theorem 2.1.** *Let  $\eta_0, u_0 \in \mathbb{H}^{s+1}(\mathbb{R})$ ,  $s > \frac{1}{2}$  real. Then there exists a positive constant  $T_0 > 0$  and a unique solution  $(\eta, u) \in C(0, T_0, \mathbb{H}^s(\mathbb{R}))$  of the Cauchy problem associated with (1.1)-(1.2) under the conditions C1, C2, C3, C4 or C5.*

*Proof.* The system of equations (1.1)-(1.2) can be written

$$(2.1) \quad U_t = AU + DF(U)$$

where

$$U = \begin{pmatrix} \eta \\ u \end{pmatrix}, \quad F(U) = \begin{pmatrix} f_1(\eta, u) \\ f_2(\eta, u) \end{pmatrix},$$

$$A = \begin{pmatrix} 0 & -(1 - b \frac{\partial^2}{\partial x^2} + b_2 \frac{\partial^4}{\partial x^4})^{-1} (1 + a \frac{\partial^2}{\partial x^2}) \frac{\partial}{\partial x} \\ -(1 - d \frac{\partial^2}{\partial x^2} + d_2 \frac{\partial^4}{\partial x^4})^{-1} (1 + a \frac{\partial^2}{\partial x^2}) \frac{\partial}{\partial x} & 0 \end{pmatrix}$$

and

$$D = \begin{pmatrix} -(1 - b \frac{\partial^2}{\partial x^2} + b_2 \frac{\partial^4}{\partial x^4})^{-1} (1 - b_2 \frac{\partial^2}{\partial x^2}) \frac{\partial}{\partial x} & 0 \\ 0 & -(1 - d \frac{\partial^2}{\partial x^2} + d_2 \frac{\partial^4}{\partial x^4})^{-1} (1 - d_2 \frac{\partial^2}{\partial x^2}) \frac{\partial}{\partial x} \end{pmatrix}.$$

The operator  $A$  has a matrix valued symbol  $\hat{A}(\xi)$ ,  $\xi \in \mathbb{R}$ , which is diagonalizable, and we note  $\tilde{A}$  the diagonalized matrix symbol of  $\hat{A}$ . Hence we can write  $\tilde{A} =$

## GLOBAL EXISTENCE

$\widehat{P}\widehat{A}(\widehat{P})^{-1}$ , where  $\widehat{P} = (\widehat{p}_{ij})$ ,  $i, j = 1, 2$  is the matrix of the eigenvectors associated with the eigenvalues  $i\lambda_1$  and  $i\lambda_2$  with

$$\lambda_1(\xi) = \frac{\xi(1 - a\xi^2)}{\sqrt{(1 + b\xi^2 + b_2\xi^4)(1 + d\xi^2 + d_2\xi^4)}} = -\lambda_2(\xi).$$

Let

$$S(t)(\varphi_1, \varphi_2) = \mathcal{F}^{-1}(e^{\tilde{A}t}(\varphi_1, \varphi_2)) = (\mathcal{F}^{-1}(e^{\tilde{A}t}\varphi_1), \mathcal{F}^{-1}(e^{\tilde{A}t}\varphi_2)) = (S_1(t)\varphi_1, S_2(t)\varphi_2),$$

be the semi-group that occurs in the computation of the ordinary differential equation  $U_t = AU$  whenever it is done in the eigenvectors basis, with for  $j = 1, 2$ ,  $S_j(t)\varphi_j = \frac{1}{2\pi} \int_{\mathbf{R}} e^{ix\xi + it\lambda_j(\xi)} \widehat{\varphi_j} d\xi$ . Then, after a few computations in which we diagonalize the matrix symbol of  $A$  and use the Duhamel formula, we obtain the following integral form of the solution to the I.V.P. (1.1)-(1.2):

$$(2.2) \quad \begin{aligned} \eta(x, t) &= p_{11} * [S_1(t)(l_{11} * \eta_0 + l_{12} * u_0) + S_2(t)(l_{11} * \eta_0 + l_{12} * u_0)] \\ &\quad + \int_0^t p_{11} * S_1(t - \tau)(h_{11} * f_1(\eta, u) + h_{12} * f_2(\eta, u)) d\tau \\ &\quad + \int_0^t p_{11} * S_2(t - \tau)(h_{21} * f_1(\eta, u) + h_{22} * f_2(\eta, u)) d\tau, \end{aligned}$$

$$(2.3) \quad \begin{aligned} u(x, t) &= p_{21} * [S_1(t)(l_{11} * \eta_0 + l_{12} * u_0) - S_2(t)(l_{11} * \eta_0 + l_{12} * u_0)] \\ &\quad + \int_0^t p_{21} * S_1(t - \tau)(h_{11} * f_1(\eta, u) + h_{12} * f_2(\eta, u)) d\tau \\ &\quad - \int_0^t p_{21} * S_2(t - \tau)(h_{21} * f_1(\eta, u) + h_{22} * f_2(\eta, u)) d\tau, \end{aligned}$$

where

$$\begin{aligned} \widehat{p}_{11} &= \widehat{p}_{12} = \frac{1}{\sqrt{1 + b\xi^2 + b_2\xi^4}}, \quad \widehat{p}_{21} = -\widehat{p}_{22} = -\frac{1}{\sqrt{1 + d\xi^2 + d_2\xi^4}}, \\ \widehat{l}_{11} &= \widehat{l}_{21} = \frac{1}{2}\sqrt{1 + b\xi^2 + b_2\xi^4}, \quad \widehat{l}_{12} = \widehat{l}_{22} = -\frac{1}{2}\sqrt{1 + d\xi^2 + d_2\xi^4}, \\ \widehat{h}_{11} &= \widehat{h}_{21} = \frac{-i\xi(1 + b_2\xi^2)}{2\sqrt{1 + b\xi^2 + b_2\xi^4}}, \quad \widehat{h}_{12} = \widehat{h}_{22} = \frac{-i\xi(1 + d_2\xi^2)}{2\sqrt{1 + d\xi^2 + d_2\xi^4}}. \end{aligned}$$

To finish the proof of the theorem 2.1 and for the sequel we need the following inequality obtained thanks to the use of the Sobolev embedding  $\mathbf{H}^s(\mathbf{R}) \subset \mathbf{L}^\infty(\mathbf{R})$ ,  $s > \frac{1}{2}$ , and thanks to the fact that  $\mathbf{H}^s(\mathbf{R})$ ,  $s > \frac{1}{2}$ , is an algebra :

$$(2.4) \quad \begin{aligned} \|f_1(\eta, u)\|_s + \|f_2(\eta, u)\|_s &\leq c(\|\eta^\alpha u\|_s + \|u^\alpha \eta\|_s + \|\eta^{\alpha+1}\|_s + \|u^{\alpha+1}\|_s) \\ &\leq c(|\eta|_\infty^{\alpha-1} \|\eta\|_s \|u\|_s + |u|_\infty^{\alpha-1} \|\eta\|_s \|u\|_s \\ &\quad + |\eta|_\infty^{\alpha-1} \|\eta\|_s^2 + |u|_\infty^{\alpha-1} \|u\|_s^2) \\ &\leq c(\|\eta\|_s^\alpha \|u\|_s + \|u\|_s^\alpha \|\eta\|_s + \|\eta\|_s^{\alpha+1} + \|u\|_s^{\alpha+1}) \end{aligned}$$

and we need the inequalities in the following lemma :

**Lemma 2.2.** *Let  $a, b, d, b_2, d_2$  in (1.1) satisfy the conditions (C1), (C2), (C3), (C4) or (C5). Then let  $\psi \in \mathbb{H}^{s+1}(\mathbb{R})$ ,  $s > \frac{1}{2}$ ,  $\mathcal{M}_1 = 1 - b \frac{\partial^2}{\partial x^2} + b_2 \frac{\partial^4}{\partial x^4}$ ,  $\mathcal{M}_2 = 1 -$*

DÉ GODEFROY AKMEL

$d \frac{\partial^2}{\partial x^2} + d_2 \frac{\partial^4}{\partial x^4}$ , and  $p_{ij}, h_{ij}, l_{ij} \forall i, j = 1, 2$  defined above. Then we have the following inequalities  $\forall i, j = 1, 2$ ,

$$(2.5) \quad \|S_j(r)\psi\|_s + \|p_{ij} * S_j(r)\psi\|_s + \|p_{ij} * S_j(r)(h_{ij} * \psi)\|_s \leq \|\psi\|_s$$

$$(2.6) \quad \|S_j(r)(h_{ij} * \psi)\|_s + \|p_{ij} * S_j(r)(l_{ij} * \psi)\|_s \leq \|\psi\|_{s+1}$$

$$(2.7) \quad \|S_j(r)(l_{ij} * \psi)\|_s \leq \|\mathcal{M}_j\psi\|_s$$

*Proof.* . *Proof of the lemma 2.2* Thanks to the definitions of  $p_{ij}, h_{ij}, l_{ij}$  and  $\mathcal{M}_j$  above, we have for all  $i, j = 1, 2$  and for all  $a, b, d, b_2, d_2$  satisfying the conditions (C1), (C2), (C3), (C4) or (C5), the inequalities

$$(2.8) \quad |\hat{p}_{ij}| + |\hat{p}_{ij}\hat{h}_{ij}| \leq c$$

$$(2.9) \quad |\hat{h}_{ij}| + |\hat{p}_{ij}\hat{l}_{ij}| \leq c(1 + \xi^2)^{\frac{1}{2}}$$

$$(2.10) \quad |\hat{l}_{11}\hat{\psi}| = |\hat{l}_{21}\hat{\psi}| = |\hat{\Lambda}_1\mathcal{F}(\mathcal{M}_1\psi)| \leq |\widehat{\mathcal{M}_1\psi}|$$

$$(2.11) \quad |\hat{l}_{12}\hat{\psi}| = |\hat{l}_{22}\hat{\psi}| = |\hat{\Lambda}_2\mathcal{F}(\mathcal{M}_2\psi)| \leq |\widehat{\mathcal{M}_2\psi}|$$

where  $\hat{\Lambda}_1(\xi) = \frac{1}{2\sqrt{1+b\xi^2+b_2\xi^4}}$  and  $\hat{\Lambda}_2(\xi) = \frac{1}{2\sqrt{1+d\xi^2+d_2\xi^4}}$ . Then with the inequalities (2.8), (2.9), (2.10), (2.11) and thanks to the Plancherel theorem and the definition of the norm  $\mathbb{H}^s$  and of  $S_j$ , we are lead to the inequalities of the lemma 2.2.  $\square$

Now, consider the complete metric space

$$F = \{(\eta, u) \in (\mathcal{C}(0, T; \mathbb{H}^s(\mathbb{R})))^2, \sup_{[0, T]} \|\eta\|_s + \sup_{[0, T]} \|u\|_s \leq \mu\}, \quad s > \frac{1}{2},$$

where  $\mu$  is a positive real constant. Then an application of the contraction-mapping-principle to  $F$  combined with the inequalities (2.4), (2.5), (2.6), (2.7), above, and an appropriate choice of  $T$  yields to the local existence result of the theorem 2.1.  $\square$

### 3. LINEAR ESTIMATES

The purpose of this section is to study the linear equation associated with the I.V.P (1.1)-(1.2) under the restrictions C1, C2, C3, or C4. We establish also linear estimates needed for the next section. For that we give some decay estimates and useful inequalities of the solution of the linearized system (1.1)-(1.2) via decay estimates of the semi-group  $S(t)(\varphi_1, \varphi_2) = (S_1(t)\varphi_1, S_2(t)\varphi_2)$ .

Consider the linear problem associated to (1.1)-(1.2):

$$(LP) \begin{cases} \eta_t - b\eta_{xxt} + b_2\eta_{xxxxx} + u_x + au_{xxx} = 0 & x, t \in \mathbb{R}, \\ u_t - du_{xxt} + d_2u_{xxxxx} + \eta_x + a\eta_{xxx} = 0 \end{cases}$$

with initial data  $\eta(x, 0) = \eta_0(x)$ ,  $u(x, 0) = u_0(x)$ . We prove the following theorem

**Theorem 3.1.** *Let  $\eta_0(x)$ ,  $J^2\eta_0(x)$ ,  $\mathcal{M}_j\eta_0(x)$ ,  $u_0(x)$ ,  $J^2u_0(x)$ ,  $\mathcal{M}_ju_0(x) \in \mathbb{H}^6(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$  where for each  $j = 1, 2$ ,  $\mathcal{M}_j$  is defined above in lemma 2.2. Then the solution  $(\eta, u)$  of the linear problem (LP) satisfies*

$$(3.1) \quad |\eta(x, t)|_{L^\infty(\mathbb{R})} + |u(x, t)|_{L^\infty(\mathbb{R})} \leq \begin{cases} c(1+t)^{-\frac{1}{4}}, & \text{in the case } C1-1, C2-1, C4-1; \\ c(1+t)^{-\frac{1}{8}}, & \text{in the case } C1-2, C2-2, C3, C4-2. \end{cases}$$

for all  $t \geq 0$ ,  $x \in \mathbb{R}$ , where  $c$  does not depend on  $x$  or  $t$ .

## GLOBAL EXISTENCE

This theorem is a consequence of the following lemma :

**Lemma 3.2.** *Let for  $j = 1, 2$ ,  $\varphi_j, J^2\varphi_j, \mathcal{M}_j\varphi_j \in \mathbb{H}^6(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$  where  $\mathcal{M}_j$  are defined above in lemma 2.2, and let  $S_j(t)$  be as defined above as the components of the semi-group  $S(t)$  that occurs in the computation of the linear equation (LP). Then for each  $i, j = 1, 2$  with  $p_{i,j}, h_{i,j}, l_{i,j}$  defined above we have the estimates,  $\forall t \geq 0, \forall x \in \mathbb{R}$ ,*

$$(3.2) \quad |p_{i,j} * S_j(t)\varphi_j(x)|_{L^\infty} + |S_j(t)\varphi_j(x)|_{L^\infty} \leq c(1+t)^{-\frac{1}{4}}(\|\varphi_j\|_6 + |\varphi_j|_1) \\ \text{if } a, b, d, b_2, d_2 \text{ satisfy } (C1-1), (C2-1) \text{ or } (C4-1),$$

$$(3.3) \quad |p_{i,j} * S_j(t)\varphi_j(x)|_{L^\infty} + |S_j(t)\varphi_j(x)|_{L^\infty} \leq c(1+t)^{-\frac{1}{8}}(\|\varphi_j\|_6 + |\varphi_j|_1) \\ \text{if } a, b, d, b_2, d_2 \text{ satisfy } (C1-2), (C2-2), (C3) \text{ or } (C4-2),$$

$$(3.4) \quad |p_{i,j} * S_j(t)(h_{i,j} * \varphi_j)|_{L^\infty} \leq c(1+t)^{-\frac{1}{4}}(\|\varphi_j\|_8 + |J^2\varphi_j|_1) \\ \text{if } a, b, d, b_2, d_2 \text{ satisfy } (C1-1), (C2-1) \text{ or } (C4-1),$$

$$(3.5) \quad |p_{i,j} * S_j(t)(h_{i,j} * \varphi_j)|_{L^\infty} \leq c(1+t)^{-\frac{1}{8}}(\|\varphi_j\|_8 + |J^2\varphi_j|_1) \\ \text{if } a, b, d, b_2, d_2 \text{ satisfy } (C1-2), (C2-2), (C3) \text{ or } (C4-2),$$

$$(3.6) \quad |p_{i,j} * S_j(t)(l_{i,j} * \varphi_j)|_{L^\infty} \leq c(1+t)^{-\frac{1}{4}}(\|\mathcal{M}_j\varphi_j\|_6 + |\mathcal{M}_j\varphi_j|_1) \\ \text{if } a, b, d, b_2, d_2 \text{ satisfy } (C1-1), (C2-1) \text{ or } (C4-1),$$

$$(3.7) \quad |p_{i,j} * S_j(t)(l_{i,j} * \varphi_j)|_{L^\infty} \leq c(1+t)^{-\frac{1}{8}}(\|\mathcal{M}_j\varphi_j\|_6 + |\mathcal{M}_j\varphi_j|_1) \\ \text{if } a, b, d, b_2, d_2 \text{ satisfy } (C1-2), (C2-2) (C3), \text{ or } (C4-2),$$

where  $c$  is independent of  $\varphi_j, x$

In order to prove the lemma 3.2 we need the following proposition.

**Proposition 3.3.** *Given  $x \in \mathbb{R}$  and  $t \in \mathbb{R}_+$ , consider the phases functions*

$$\psi_j(\xi) = \lambda_j(\xi) + t^{-1}x\xi, \quad j = 1, 2$$

where  $\lambda_j$  are defined above. Then under the conditions  $C1, C2, C3$  or  $C4$ ,  $\psi_j, j = 1, 2$  have a finite number of stationary points. Moreover for each  $j = 1, 2$ , there exists at least a stationary point  $\xi_{sj}$  of  $\psi_j$  which verifies

$$(3.8) \quad R(\xi_{sj}^2) \geq 0$$

where

$$R(y) = 1 - 3ay - [bd + b_2 + d_2 + 2a(b+d)]y^2 - [2(b_2d + bd_2) + a(bd + b_2 + d_2)]y^3 - 3b_2d_2y^4 + ab_2d_2y^5$$

is variable whenever  $a, b, d, b_2, d_2$  satisfy the conditions  $C1, C2, C3$  or  $C4$

*Proof.* Proof of proposition 3.3. As given above we have (for the full system (1.1)-(1.2))

$$\lambda_j(\xi) = (-1)^{j-1} \frac{\xi(1 - a\xi^2)}{\sqrt{(1 + b\xi^2 + b_2\xi^4)(1 + d\xi^2 + d_2\xi^4)}}.$$



Then for  $j = 1, 2$ ,

$$\psi_j'(\xi) = 0 \Leftrightarrow \lambda_j'(\xi) + t^{-1}x\xi = 0$$

$$\Leftrightarrow \{1 - 3a\xi^2 - [bd + b_2 + d_2 + 2a(b + d)]\xi^4 - [2(b_2d + bd_2) + a(bd + b_2 + d_2)]\xi^6 - 3b_2d_2\xi^8 + ab_2d_2\xi^{10}\}^2 - x^2t^{-2}(1 + b\xi^2 + b_2\xi^4)^3(1 + d\xi^2 + d_2\xi^4)^3 = 0$$

so that

$$(3.9) \quad \psi_j'(\xi) = 0 \Leftrightarrow P(\xi^2) = 0$$

where

$$P(y) = R(y)^2 - x^2t^{-2}(1 + by + b_2y^2)^3(1 + dy + d_2y^2)^3$$

with

$$R(y) = 1 - 3ay - [bd + b_2 + d_2 + 2a(b + d)]y^2 - [2(b_2d + bd_2) + a(bd + b_2 + d_2)]y^3 - 3b_2d_2y^4 + ab_2d_2y^5.$$

Therefore thanks to (3.9), the stationary points of  $\psi$  are such that their square are roots of  $P(y)$ . Hence since  $P(y)$  is a polynomial of degree 12 so that it has at most 12 roots, we deduce from (3.9) that each  $\psi_j$ ,  $j = 1, 2$  has at most 24 stationary points in  $\mathbb{R}$ . Let us prove now the second part of the proposition 3.3 (for all the cases  $C1, C2, C3$  or  $C4$ ).

*Proof for the case C1.*

If  $a, b, d, b_2, d_2$  satisfy  $C1$ , then (3.9) is verified by

$$P(y) = R(y)^2 - x^2t^{-2}(1 + by)^3(1 + dy + d_2y^2)^3$$

where

$$R(y) = 1 - 3ay - (bd + d_2 + 2ab + 2ad)y^2 - (2bd_2 + abd + ad_2)y^3.$$

We see that for fixed  $x$  and  $t$  large enough,

$$(3.10) \quad P(0) = 1 - x^2t^{-2} \geq 0.$$

Moreover,  $R(0) = 1 > 0$  and  $\lim_{\xi \rightarrow \infty} R(\xi^2) = -\infty$  so that since  $\xi \mapsto R(\xi^2)$  is continuous, there exists at least a  $\xi_0 \geq 0$  such that  $R(\xi_0^2) = 0$ . Henceforth, with a such  $\xi_0$  we have

$$(3.11) \quad P(\xi_0^2) = -x^2t^{-2}(1 + b\xi_0^2)^3(1 + d\xi_0^2 + d_2\xi_0^4)^3.$$

Therefore thanks to (3.11) and since  $b, d, d_2$  are positives, we have

$$(3.12) \quad P(\xi_0^2) \leq 0.$$

Hence since  $\xi \mapsto P(\xi^2)$  is continuous, we deduce from (3.10)-(3.12) that the equation  $P(\xi^2) = 0$  has at least one solution  $\xi_s \in ]0, \xi_0[$  where  $\xi_0 \geq 0$  is a root of  $R(\xi^2)$ . We deduce with (3.9) that for each  $j = 1, 2$ ,  $\psi_j$  has at least one stationary point  $\xi_{sj} \in ]0, \xi_0[$ . Moreover, since for each  $j = 1, 2$ ,  $\psi_j$  has a stationary point  $\xi_{sj}$  which verifies

$$(3.13) \quad 0 < \xi_{sj} < \xi_0$$

where  $\xi_0$  is a root of  $R(\xi^2)$ , and hence since for all  $y \geq 0$ ,  $R(y)$  is a decreasing function, we get for each  $j = 1, 2$ , thanks to (3.13)

$$(3.14) \quad R(\xi_{sj}^2) \geq R(\xi_0^2) = 0$$

## GLOBAL EXISTENCE

and the inequality (3.14) finishes the proof of the proposition 3.3 for the case  $C1$ . (Note here that if we had  $\xi_0 < \xi_{sj} < 0$  then we would have  $0 < \xi_{sj}^2 < \xi_0^2$  and consequently (3.14)).

*Proof for the case C2.*

The proof of the proposition 3.3 for the case C2 follows from that of the case C1 by symetry.

*Proof for the cases C3 and C4.*

For the cases  $C3$  and  $C4$ , we see that the functions  $P(y)$  and  $R(y)$  from the case  $C3$  and that from the case  $C4$  verify the same properties as that of the functions  $P(y)$  and  $R(y)$  in the proof for the case  $C1$ , particularly the properties (3.9), (3.10), (3.12), (3.14). Therefore, following the same lines of the proof for the case  $C1$ , we find the proof of the second part of the proposition 3.3 for the cases  $C3$  and  $C4$ . This finishes the proof of the proposition 3.3.  $\square$

Let us prove now the lemma 3.2

*Proof. . Proof of the lemma 3.2*

If  $0 \leq t \leq 1$ , we have thanks to the definitions of  $S_j(t)$ ,  $j = 1, 2$ , above in section 2, and  $\psi_j$  in proposition 3.3, and thanks to the Schwartz inequality, we have

$$(3.15) \quad |S_j(t)\varphi_j(x)| = \frac{1}{2\pi} \left| \int_{\mathbf{R}} e^{it\psi_j(\xi)} \hat{\varphi}_j(\xi) d\xi \right| \leq c \left( \int_{\mathbf{R}} (1 + \xi^2)^{-6} d\xi \right)^{\frac{1}{2}} \|\varphi_j\|_6 \\ \leq c \|\varphi_j\|_6 \leq c(1+t)^{-\frac{1}{4}} \|\varphi_j\|_6.$$

If  $t \geq 1$ , let  $\Omega = \{\xi \in \mathbf{R} \mid |\xi| \leq t^{\frac{1}{m}}\}$ ,  $m \geq 1$ , and  $q_{jt}(\xi) = \chi_{\Omega}(\xi) e^{it\lambda_j(\xi)}$ ; then thanks to the Schwartz and the Young inequalities,

$$(3.16) \quad |S_j(t)\varphi_j(x)| = \frac{1}{2\pi} \left| \left( \int_{\Omega} + \int_{\Omega^c} \right) e^{it\lambda_j(\xi) + ix\xi} \hat{\varphi}_j(\xi) d\xi \right| \\ \leq c |\check{q}_{jt}(x) * \varphi_j(x)|_{\infty} \\ + c \left( \int_{\Omega^c} (1 + \xi^2)^{-6} d\xi \right)^{\frac{1}{2}} \left( \int_{\Omega^c} (1 + \xi^2)^6 |\hat{\varphi}_j(\xi)|^2 d\xi \right)^{\frac{1}{2}} \\ \leq c |\check{q}_{jt}(x)|_{\infty} \|\varphi_j\|_1 + ct^{-\frac{1}{m}} \|\varphi_j\|_6.$$

It remains to estimate  $\check{q}_{jt}(x)$ . We need for the sequel the following notations: Let  $\mathcal{E}_s = \{\xi \in \mathbf{R}, \psi'_j(\xi) = 0, j = 1, 2\}$  be the set of stationary points of  $\psi_j$ . We know from the proposition 3.3 that  $\mathcal{E}_s$  has a finite number of elements. Hence set for  $m \geq 1$ ,

$$\mathcal{N}(\zeta, t^{-\frac{1}{m}}) = \bigcup_{\zeta \in \mathcal{E}_s} \bar{B}(\zeta, t^{-\frac{1}{m}}) \bigcup_{\zeta \in \mathcal{E}_s} \{\xi \in \mathbf{R} \mid |\xi + \zeta| \leq t^{-\frac{1}{m}}\} \bigcup \{\xi \in \mathbf{R} \mid |\xi| \leq t^{-\frac{1}{m}}\}$$

where for each  $\zeta \in \mathcal{E}_s$ ,  $\bar{B}(\zeta, t^{-\frac{1}{m}}) = \{\xi \in \mathbf{R} \mid |\xi - \zeta| \leq t^{-\frac{1}{m}}\}$ . Then we have

$$(3.17) \quad \check{q}_{jt}(x) = \left( \int_{\Omega \cap \mathcal{N}(\zeta, t^{-\frac{1}{m}})} + \int_{\Omega \cap \{\mathcal{N}(\zeta, t^{-\frac{1}{m}})\}^c} \right) e^{it\lambda_j(\xi) + ix\xi} d\xi = I_1 + I_2.$$

Since from proposition 3.3,  $\text{card}(\mathcal{E}_s) < \infty$ , we get

$$(3.18) \quad |I_1| \leq \int_{\Omega \cap \mathcal{N}(\zeta, t^{-\frac{1}{m}})} d\xi \leq \sum_{\zeta \in \mathcal{E}_s} \int_{\bar{B}(\zeta, t^{-\frac{1}{m}})} d\xi \\ + \sum_{\zeta \in \mathcal{E}_s} \int_{\{|\xi + \zeta| \leq t^{-\frac{1}{m}}\}} d\xi + \int_{\{|\xi| \leq t^{-\frac{1}{m}}\}} d\xi \leq ct^{-\frac{1}{m}}.$$

For  $I_2$ , we point out that on  $\{\mathcal{N}(\zeta, t^{-\frac{1}{m}})\}^c$ ,  $\psi_j$ ,  $j = 1, 2$  have no stationary point so that we can integrate  $I_2$  by parts as follows:

$$(3.19) \quad |I_2| = \left| \int_{\Omega \cap \{\mathcal{N}(\zeta, t^{-\frac{1}{m}})\}^c} e^{it\psi_j(\xi)} d\xi \right| \\ = t^{-1} \left| \int_{\Omega \cap \{\mathcal{N}(\zeta, t^{-\frac{1}{m}})\}^c} \frac{1}{\frac{\partial}{\partial \xi} \psi_j(\xi)} \frac{\partial}{\partial \xi} \left( e^{it\psi_j(\xi)} \right) d\xi \right| \\ \leq t^{-1} \int_{\Omega \cap \{\mathcal{N}(\zeta, t^{-\frac{1}{m}})\}^c} \left| \frac{\partial}{\partial \xi} \left( \frac{1}{\frac{\partial}{\partial \xi} \psi_j(\xi)} \right) \right| d\xi \\ + t^{-1} \int_{\partial\{\Omega \cap \{\mathcal{N}(\zeta, t^{-\frac{1}{m}})\}^c\}} \frac{d\xi}{\left| \frac{\partial}{\partial \xi} \psi_j(\xi) \right|} \\ \leq ct^{-1} \int_{\Omega \cap \{\mathcal{N}(\xi_{sj}, t^{-\frac{1}{m}})\}^c} \frac{\left| \frac{\partial^2}{\partial \xi^2} \psi_j(\xi) \right|}{\left| \frac{\partial}{\partial \xi} \psi_j(\xi) \right|^2} d\xi$$

where for each  $j = 1, 2$ ,  $\xi_{sj}$  is a stationary point of  $\psi_j$ , which satisfies the property  $R(\xi_{sj}^2) \geq 0$  in the proposition 3.3.

Before continue, let us give the needed following inequalities useful for the proof of the lemma 3.2 : We claim that on  $\Omega \cap \{\mathcal{N}(\xi_{sj}, t^{-\frac{1}{m}})\}^c$ , and for  $m \geq 3$ , if  $a, b, d, b_2, d_2$  satisfy (C1-1), (C2-1) or (C4-1), then for each  $j = 1, 2$ ,

$$(3.20) \quad \frac{|\psi_j''(\xi)|}{|\psi_j'(\xi)|^2} \leq \begin{cases} ct^{\frac{3}{m}} & \text{if } |\xi| \leq 1 \\ ct^{\frac{14}{5m}} & \text{otherwise} \end{cases}$$

and if  $a, b, d, b_2, d_2$  satisfy (C1-2), (C2-2), (C3) or (C4-2), then for each  $j = 1, 2$ ,

$$(3.21) \quad \frac{|\psi_j''(\xi)|}{|\psi_j'(\xi)|^2} \leq \begin{cases} ct^{\frac{7}{m}} & \text{if } |\xi| \leq 1 \\ ct^{\frac{3}{m}} & \text{otherwise.} \end{cases}$$

*Proof.* Proof of (3.20) and (3.21).

Thanks to  $\psi_j$  and  $\lambda_j$  as defined above in proposition 3.3 and in section 2, we have :

$$\psi_j'(\xi) = \lambda_j'(\xi) + t^{-1}x.$$

We find (after obvious computations),

$$\lambda_j'(\xi) = (-1)^{j-1} \frac{R(\xi^2)}{\sqrt{Q^3(\xi^2)}}$$

where for the full system (1.1)-(1.2),

$$R(y) = \{1 - 3ay - (bd + b_2 + d_2 + 2ab + 2ad)y^2 - (2b_2d + 2bd_2 + abd + ab_2 + 2ad_2)y^3 - 3b_2d_2y^4 + ab_2d_2y^5\}$$

and

$$Q(y) = (1 + by + b_2y^2)(1 + dy + d_2y^2).$$

## GLOBAL EXISTENCE

Then, with  $\xi_{sj} \in \mathcal{E}_s$  chosen as in the proposition 3.3, we get

$$(3.22) \quad |\psi'_j(\xi)| = |\psi'_j(\xi_{sj}) - \psi'_j(\xi)| = |\lambda'_j(\xi_{sj}) - \lambda'_j(\xi)|$$

where

$$(3.23) \quad \begin{aligned} \lambda'_j(\xi_{sj}) - \lambda'_j(\xi) &= (-1)^{j-1} \left\{ \frac{R(\xi_{sj}^2)}{\sqrt{Q^3(\xi_{sj}^2)}} - \frac{R(\xi^2)}{\sqrt{Q^3(\xi^2)}} \right\} \\ &= \frac{(-1)^{j-1}}{\sqrt{Q^3(\xi_{sj}^2)Q^3(\xi^2)}} \left\{ \frac{R(\xi_{sj}^2)(Q(\xi^2) - Q(\xi_{sj}^2))(Q^2(\xi^2) + Q(\xi^2)Q(\xi_{sj}^2) + Q^2(\xi_{sj}^2))}{\sqrt{Q^3(\xi_{sj}^2) + \sqrt{Q^3(\xi^2)}}} \right. \\ &\quad \left. + \sqrt{Q^3(\xi_{sj}^2)} (R(\xi_{sj}^2) - R(\xi^2)) \right\}. \end{aligned}$$

The following relations are also useful for the proof of the inequalities (3.20), (3.21):

$$(3.24) \quad |\xi_1| + |\xi_2| = |\xi_1 + \xi_2| \quad \text{if } \text{sgn}(\xi_1) = \text{sgn}(\xi_2)$$

$$(3.25) \quad |\xi_1| + |\xi_2| = |\xi_1 - \xi_2| \quad \text{if } \text{sgn}(\xi_1) \neq \text{sgn}(\xi_2)$$

The proof of (3.24) and (3.25) is as follows : If  $\text{sgn}(\xi_1) = \text{sgn}(\xi_2)$  then

$$|\xi_1| + |\xi_2| = ||\xi_1| + |\xi_2|| = |\text{sgn}(\xi_1)\xi_1 + \text{sgn}(\xi_2)\xi_2| = |\xi_1 + \xi_2|$$

If  $\text{sgn}(\xi_1) \neq \text{sgn}(\xi_2)$  that is if  $\text{sgn}(\xi_1) = -\text{sgn}(\xi_2)$  then

$$|\xi_1| + |\xi_2| = ||\xi_1| + |\xi_2|| = |\text{sgn}(\xi_1)\xi_1 + \text{sgn}(\xi_2)\xi_2| = |\xi_1 - \xi_2|.$$

We are now able to pull up the proof of the inequalities (3.20) and (3.21) for the different cases C1, C2, C3, C4 :

*Proof of (3.20) for the case C1-1.*

In the case C1-1 we have thanks to the definitions of  $R$  and  $Q$  above,

$$R(\xi_{sj}^2) - R(\xi^2) = (\xi^2 - \xi_{sj}^2)[3a + (bd + d_2 + 2ab + 2ad)(\xi^2 + \xi_{sj}^2) + (2bd_2 + ad_2 + abd)(\xi^{24} + \xi^2\xi_{sj}^2 + \xi_{sj}^4)]$$

and

$$Q(\xi^2) - Q(\xi_{sj}^2) = (\xi^2 - \xi_{sj}^2)[b + d + (bd + d_2)(\xi^2 + \xi_{sj}^2) + bd_2(\xi^{24} + \xi^2\xi_{sj}^2 + \xi_{sj}^4)]$$

so that thanks to (3.23),

$$(3.26) \quad \begin{aligned} \lambda'_j(\xi_{sj}) - \lambda'_j(\xi) &= \frac{(-1)^{j-1}(\xi^2 - \xi_{sj}^2)}{\sqrt{Q^3(\xi_{sj}^2)Q^3(\xi^2)}} \left\{ \frac{R(\xi_{sj}^2)(Q^2(\xi^2) + Q(\xi^2)Q(\xi_{sj}^2) + Q^2(\xi_{sj}^2))}{\sqrt{Q^3(\xi_{sj}^2) + \sqrt{Q^3(\xi^2)}}} [b + d(bd + d_2)(\xi^2 + \xi_{sj}^2) \right. \\ &\quad \left. + bd_2(\xi^{24} + \xi^2\xi_{sj}^2 + \xi_{sj}^4)] + \sqrt{Q^3(\xi_{sj}^2)} [3a + (bd + d_2 + 2ab + 2ad)(\xi^2 + \xi_{sj}^2) \right. \\ &\quad \left. + (2bd_2 + ad_2 + abd)(\xi^{24} + \xi^2\xi_{sj}^2 + \xi_{sj}^4)] \right\}. \end{aligned}$$

DÉ GODEFROY AKMEL

Then, since by proposition 3.3  $R(\xi_{sj}^2) \geq 0$ , and since  $Q(\xi^2) \geq 0 \forall \xi \in \mathbf{R}$ , we find with (3.22), (3.26), and the fact that  $a, b, d, b_2, d_2 \geq 0$ ,

$$\begin{aligned} |\psi'_j(\xi)| &\geq \frac{|\xi^2 - \xi_{sj}^2|}{\sqrt{Q^3(\xi^2)}} \{3a + (bd + d_2 + 2ab + 2ad)\xi^2 + (2bd_2 + ad_2 + abd)\xi^4\} \\ (3.27) \quad &\geq \frac{c|\xi^2 - \xi_{sj}^2|}{(1 + b\xi^2)^{\frac{3}{2}}(1 + d\xi^2 + d_2\xi^4)^{\frac{1}{2}}}. \end{aligned}$$

On the other hand, we find here with obvious computations,

$$(3.28) \quad |\psi''_j(\xi)| \leq \frac{c|\xi|}{(1 + b\xi^2)^{\frac{3}{2}}(1 + d\xi^2 + d_2\xi^4)^{\frac{1}{2}}}.$$

Then with (3.27), (3.28), we find

$$(3.29) \quad \frac{|\psi''_j(\xi)|}{|\psi'_j(\xi)|^2} \leq \frac{c|\xi|(1 + b\xi^2)^{\frac{3}{2}}(1 + d\xi^2 + d_2\xi^4)^{\frac{1}{2}}}{|\xi^2 - \xi_{sj}^2|^2}.$$

Hence if  $|\xi| \geq 1$  we have on  $\Omega \cap \{\mathcal{N}(\xi_{sj}, t^{-\frac{1}{m}})\}^c$  thanks to (3.29),

$$(3.30) \quad \frac{|\psi''_j(\xi)|}{|\psi'_j(\xi)|^2} \leq \frac{c|\xi|^6}{|\xi - \xi_{sj}|^2|\xi + \xi_{sj}|^2}$$

so that when  $\text{sgn}(\xi) = \text{sgn}(\xi_{sj})$  we get on  $\Omega \cap \{\mathcal{N}(\xi_{sj}, t^{-\frac{1}{m}})\}^c$ , thanks to (3.30) and (3.24),

$$\begin{aligned} \frac{|\psi''_j(\xi)|}{|\psi'_j(\xi)|^2} &\leq \frac{c|\xi|^6}{|\xi - \xi_{sj}|^2(|\xi| + |\xi_{sj}|)^2} \\ (3.31) \quad &\leq \frac{c|\xi|^6}{|\xi - \xi_{sj}|^2|\xi|^2} = \frac{c|\xi|^4}{|\xi - \xi_{sj}|^2} \leq ct^{\frac{4}{5m} + \frac{2}{m}} \leq ct^{\frac{14}{5m}} \end{aligned}$$

and likewise when  $\text{sgn}(\xi) \neq \text{sgn}(\xi_{sj})$  we find with (3.30) and (3.25),

$$(3.32) \quad \frac{|\psi''_j(\xi)|}{|\psi'_j(\xi)|^2} \leq \frac{c|\xi|^6}{|\xi + \xi_{sj}|^2|\xi|^2} = \frac{c|\xi|^4}{|\xi + \xi_{sj}|^2} \leq ct^{\frac{4}{5m} + \frac{2}{m}} \leq ct^{\frac{14}{5m}}.$$

Hence (3.31) and (3.32) give for  $|\xi| \geq 1$  and on  $\Omega \cap \{\mathcal{N}(\xi_{sj}, t^{-\frac{1}{m}})\}^c$ ,

$$(3.33) \quad \frac{|\psi''_j(\xi)|}{|\psi'_j(\xi)|^2} = ct^{\frac{14}{5m}}.$$

On the other hand, if  $|\xi| \leq 1$  then thanks to (3.29) and proceeding as in (3.31) and (3.32) we find on  $\Omega \cap \{\mathcal{N}(\xi_{sj}, t^{-\frac{1}{m}})\}^c$ ,

$$(3.34) \quad \frac{|\psi''_j(\xi)|}{|\psi'_j(\xi)|^2} \leq ct^{\frac{3}{m}}.$$

This finishes the proof of the inequality (3.20) for the case (C1-1).

The proof of (3.20) for the cases (C2-1) and (C4-1) and that of (3.21) for the cases (C1-2), (C2-2), C3 and (C4-2), follows the same lines as that of the proof of (3.20) for the case C1-1, using the  $R(y)$  and the  $Q(y)$  that correspond to these cases. This finishes the proof of the inequalities (3.20) and (3.21).  $\square$

## GLOBAL EXISTENCE

Now, with the inequality (3.20) in hands and thanks to (3.19), we find for the cases (C1-1), (C2-1), and (C4-1),

$$(3.35) \quad |I_2| \leq ct^{-1} \left( \int_{\{|\xi| \leq 1\}} t^{\frac{3}{m}} d\xi + \int_{\{1 \leq |\xi| \leq t^{\frac{1}{5m}}\}} t^{\frac{14}{5m}} d\xi \right) \leq ct^{-\frac{m-3}{m}}.$$

Likewise for the cases (C1-2), (C2-2), C3 and (C4-2), we find in the same ways thanks to the inequalities (3.19) and (3.21)

$$(3.36) \quad |I_2| \leq ct^{-\frac{m-7}{m}}.$$

Therefore, taking  $m = 4$ , we find for the cases (C1-1), (C2-1), (C4-1), thanks to (3.35), (3.18), (3.17), and (3.16),  $\forall t \geq 1$ ,

$$(3.37) \quad |S_j(t)\varphi_j(x)| \leq ct^{-\frac{1}{4}}(|\varphi_j|_1 + \|\varphi_j\|_6) \leq c(1+t)^{-\frac{1}{4}}(|\varphi_j|_1 + \|\varphi_j\|_6)$$

and taking  $m = 8$ , we find for the cases (C1-2), (C2-2), C3 or (C4-2), thanks to (3.36), (3.18), (3.17), and (3.16),,

$$(3.38) \quad |S_j(t)\varphi_j(x)| \leq c(1+t)^{-\frac{1}{8}}(|\varphi_j|_1 + \|\varphi_j\|_6) \quad \forall t \geq 1.$$

(3.37), (3.38), with (3.15) give a part of the inequalities (3.2) and (3.3) in the lemma 3.2.

In order to finish the proof of the lemma 3.2, we need the following lemmas which are also useful for the sequel.

**Lemma 3.4.** *Let  $h$  be such that  $(1 - \frac{\partial^2}{\partial \xi^2})\hat{h}(\xi) \in \mathbf{L}^2(\mathbf{R})$ . Then we have  $h(x) \in \mathbf{L}^1(\mathbf{R}) \cap \mathbf{L}^2(\mathbf{R})$ .*

**Lemma 3.5.** *Let for each  $i, j = 1, 2$ ,  $p_{ij}$  be defined as above in section 2. Then for  $a, b, d, b_2, d_2$  satisfying C1, C2, C3 or C4, we have  $p_{ij} \in \mathbf{L}^1(\mathbf{R}) \cap \mathbf{L}^2(\mathbf{R})$ .*

*Proof.* *Proof of lemma 3.4 and lemma 3.5.* We have thanks to the Plancherel theorem,

$$(3.39) \quad \begin{aligned} \int_{\mathbf{R}} (1+x^2)^2 |h(x)|^2 dx &= \int_{\mathbf{R}} |\mathcal{F}((1+x^2)h(x))(\xi)|^2 d\xi \\ &= \int_{\mathbf{R}} |(1 - \frac{\partial^2}{\partial \xi^2})\hat{h}(\xi)|^2 d\xi \leq c < \infty. \end{aligned}$$

This shows that  $h \in \mathbf{L}^2(\mathbf{R})$ . Moreover, with (3.39) and thanks to the Schwartz inequality, we get

$$(3.40) \quad \int_{\mathbf{R}} |h(x)| dx \leq \left( \int_{\mathbf{R}} (1+x^2)^{-2} dx \right)^{\frac{1}{2}} \left( \int_{\mathbf{R}} (1+x^2)^2 |h(x)|^2 dx \right)^{\frac{1}{2}} \leq c < \infty$$

that is  $h \in \mathbf{L}^1(\mathbf{R})$  and the lemma 3.4 is proven. Now, thanks to the lemma 3.4 the definition of  $p_{ij}$  in the section 2, the proof of the lemma 3.5 follows immediately.  $\square$

We are now able to finish the proof of the lemma 3.2. We begin by ending the proof of the inequalities (3.2) and (3.3) : Since from lemma 3.5  $p_{ij} \in \mathbf{L}^1(\mathbf{R})$   $i, j = 1, 2$ , and thanks to the estimates of  $S_j(t)$ ,  $i, j = 1, 2$ , in (3.2) and the Young inequality, we find for the cases C1-1, C2-1, C4-1,

$$(3.41) \quad \begin{aligned} |p_{ij} * S_j(t)\varphi_j(x)|_{\infty} &\leq |p_{ij}|_1 |S_j(t)\varphi_j(x)|_{\infty} \\ &\leq c(1+t)^{-\frac{1}{4}}(|\varphi_j|_1 + \|\varphi_j\|_6) \quad \forall t \geq 0, \end{aligned}$$

which finishes the proof of the inequality (3.2). We finish the proof of the inequality (3.3) in the same manner. To prove the others inequalities (3.4), (3.5), (3.6), (3.7) of the lemma 3.2, we use the definitions of  $p_{ij}, h_{ij}, l_{ij}$   $i, j = 1, 2$ , in the section 2, and we proceed as above and as in the proof of the lemma 2.2. This ends the proof of the lemma 3.2.  $\square$

*Proof. Proof of the theorem 3.1.* Writing the solution  $(\eta, u)$  of  $(LP)$  in its integral form as follows,

$$(3.42) \quad \eta(x, t) = p_{11} * [S_1(t)(l_{11} * \eta_0 + l_{12} * u_0) + S_2(t)(l_{11} * \eta_0 + l_{12} * u_0)]$$

$$(3.43) \quad u(x, t) = p_{21} * [S_1(t)(l_{11} * \eta_0 + l_{12} * u_0) - S_2(t)(l_{11} * \eta_0 + l_{12} * u_0)].$$

the proof of the theorem 3.1 follows immediately from the inequalities of the lemma 3.2.  $\square$

#### 4. GLOBAL EXISTENCE AND DECAY FOR THE SOLUTION OF THE NL SYSTEM.

Let us state now the results of the global existence and decay properties of the solution of the nonlinear system (1.1).

**Theorem 4.1.** *Let  $\alpha > 5$  and let  $\eta_0(x), J^2\eta_0(x), \mathcal{M}_j\eta_0(x), u_0(x), J^2u_0(x), \mathcal{M}_ju_0(x) \in \mathbb{H}^7(\mathbb{R}) \cap \mathbb{L}^1(\mathbb{R})$  where  $\mathcal{M}_1 = 1 - b\frac{\partial^2}{\partial x^2} + b_2\frac{\partial^4}{\partial x^4}$ ,  $\mathcal{M}_2 = 1 - d\frac{\partial^2}{\partial x^2} + d_2\frac{\partial^4}{\partial x^4}$ . Suppose that, for each  $j = 1, 2$ ,*

$$|\mathcal{M}_j\eta_0(x)|_1 + |\mathcal{M}_ju_0(x)|_1 + \|\mathcal{M}_j\eta_0(x)\|_6 + \|\mathcal{M}_ju_0(x)\|_6 < \delta,$$

*$\delta$  sufficiently small. Then if  $\alpha > 5$ , the solution  $(\eta, u)$  of the Cauchy problem associated to (1.1)-(1.2) with  $a, b, d, b_2, d_2$  satisfying the conditions (C1-1), (C2-1) or (C4-1) verifies*

$$(4.1) \quad |\eta(x, t)|_\infty + |u(x, t)|_\infty \leq c(1+t)^{-\frac{1}{4}}, \quad \forall t \geq 0,$$

$$(4.2) \quad \|\eta(x, t)\|_8 + \|u(x, t)\|_8 \leq c.$$

*Otherwise if  $\alpha > 9$ , then the solution  $(\eta, u)$  of the Cauchy problem associated to (1.1)-(1.2) with  $a, b, d, b_2, d_2$  satisfying the conditions (C1-2), (C2-2), (C3) or (C4-2) verifies*

$$(4.3) \quad |\eta(x, t)|_\infty + |u(x, t)|_\infty \leq c(1+t)^{-\frac{1}{8}}, \quad \forall t \geq 0,$$

$$(4.4) \quad \|\eta(x, t)\|_8 + \|u(x, t)\|_8 \leq c.$$

*Proof.* In addition with the inequality (2.4) and those of the lemma 2.2, we need for the proof of the theorem 4.1, the following inequalities: From the definition of  $f_1, f_2$  given in (1.1)-(1.2), we find thanks to the Holder inequalities,

$$(4.5) \quad |J^2(f_1(\eta, u))|_1 \leq c(|\eta|_\infty^{\alpha-1}\|u\|_2\|\eta\|_2 + |u|_\infty\|\eta|_\infty^{\alpha-2}\|\eta\|_2^2)$$

and

$$(4.6) \quad |J^2(f_2(\eta, u))|_1 \leq c|u|_\infty^{\alpha-1}\|u\|_2^2 + c|\eta|_\infty^{\alpha-1}\|\eta\|_2^2.$$

Now define the quantity

$$Q(t) = \sup_{0 \leq \tau \leq t} \{(1+\tau)^{\frac{1}{4}}(|\eta(\tau)|_\infty + |u(\tau)|_\infty) + \|\eta\|_8 + \|u\|_8\}$$

We will consider here the integral form of the nonlinear solution of (1.1)-(1.2) as given by (2.2)-(2.3) in the section 2 above. Therefore, taking the  $L^\infty$  norm of the

## GLOBAL EXISTENCE

solution  $(\eta, u)$  written in its integral form, and thanks to the inequalities in the lemma 3.2 of section 3, and the inequalities (4.5), (4.6), (2.4), we find for each  $j = 1, 2$ , if  $a, b, c, d, b_2, d_2$  satisfy (C1-1), (C2-1), or (C4-1),

$$\begin{aligned}
 (4.7) \quad |\eta(x, t)|_\infty + |u(x, t)|_\infty &\leq c(1+t)^{-\frac{1}{4}}(|\mathcal{M}_j \eta_0(x)|_1 + \|\mathcal{M}_j \eta_0(x)\|_6 \\
 &\quad + |\mathcal{M}_j u_0(x)|_1 + \|\mathcal{M}_j u_0(x)\|_6) \\
 &\quad + cQ(t)^{\alpha+1} \int_0^t (1+(t-\tau))^{-\frac{1}{4}} (1+\tau)^{-\frac{\alpha-1}{4}} d\tau.
 \end{aligned}$$

But since for  $\alpha > 5$ ,

$$\int_0^t (1+(t-\tau))^{-\frac{1}{4}} (1+\tau)^{-\frac{\alpha-1}{4}} d\tau \leq c(1+t)^{-\frac{1}{4}},$$

we deduce from (4.7) that for  $\alpha > 5$ , when (1.1)-(1.2) is under the conditions (C1-1), (C2-1), or (C4-1),

$$\begin{aligned}
 (4.8) \quad (1+\tau)^{\frac{1}{4}}(|\eta(\tau)|_\infty + |u(\tau)|_\infty) &\leq c(|\mathcal{M}_j \eta_0(x)|_1 + \|\mathcal{M}_j \eta_0(x)\|_6 \\
 &\quad + |\mathcal{M}_j u_0(x)|_1 + \|\mathcal{M}_j u_0(x)\|_6 \\
 &\quad + Q(t)^{\alpha+1}).
 \end{aligned}$$

Furthermore, when (1.1)-(1.2) is under the conditions (C1-1), (C2-1), (C4-1), we get for  $\alpha > 5$  thanks to (2.4) and the inequalities of the lemma in section 2 above,

$$\begin{aligned}
 (4.9) \quad \|\eta(x, t)\|_8 + \|u(x, t)\|_8 &\leq c\{\|\eta_0(x)\|_9 + \|u_0(x)\|_9 \\
 &\quad + Q(t)^{\alpha+1} \int_0^t (1+\tau)^{-\frac{\alpha-1}{4}} d\tau\} \\
 &\leq c\{\|\eta_0(x)\|_9 + \|u_0(x)\|_9 + Q(t)^{\alpha+1}\}.
 \end{aligned}$$

Then, thanks to (4.8), (4.9) and the definition of  $Q(t)$  above, we are lead for  $\alpha > 5$ , to the inequality

$$\begin{aligned}
 (4.10) \quad Q(t) &\leq c\{|\mathcal{M}_j \eta_0|_1 + |\mathcal{M}_j u_0|_1 + \|\mathcal{M}_j \eta_0\|_6 + \|\mathcal{M}_j u_0\|_6 \\
 &\quad + \|\mathcal{M}_j u_0\|_6 + \|\eta_0\|_9 + \|u_0\|_9 + Q(t)^{\alpha+1}\}.
 \end{aligned}$$

Henceforth, thanks to the inequality (4.10), we find that, if

$$(4.11) \quad |\mathcal{M}_j \eta_0|_1 + |\mathcal{M}_j u_0|_1 + \|\mathcal{M}_j \eta_0\|_6 + \|\mathcal{M}_j u_0\|_6 + \|\eta_0\|_9 + \|u_0\|_9 < \delta$$

with  $\delta$  small enough, then  $Q(t)$  is bounded. Indeed, the inequality (4.10) is satisfied if  $Q(t) \in [0, \beta_1] \cup [\beta_2, \infty[$  with  $0 < \beta_1 < \beta_2 < \infty$  (since  $\delta$  is small). Thereby, since by the definition of  $Q(t)$ ,  $Q(0) = |\eta_0|_\infty + |u_0|_\infty + \|\eta_0\|_8 + \|u_0\|_8$ , we have thanks to the Sobolev embedding,  $\mathbf{H}^8(R) \subset \mathbf{L}^\infty(R)$  and with (4.11),

$$(4.12) \quad Q(0) \leq c(\|\eta_0\|_8 + \|u_0\|_8) \leq c\delta.$$

Then, the continuity of  $Q(t)$  and the inequality (4.12) allow us to conclude that  $Q(t)$  remains bounded for  $\delta$  small enough and for all  $t \geq 0$ ; otherwise, (4.12) would be contradicted. Thus we have obtained for  $\alpha > 5$ , a bound of  $Q(t)$  and consequently, a bound and a decay estimate of the local solution which permit us to extend globally,



for  $\alpha > 5$  and for small initial values, the local solution of the system (1.1)-(1.2) under the conditions (C1-1), (C2-1) or (C4-1). On the other hand, following the same lines above, we find for  $\alpha > 9$ , the same results as above for the solution of (1.1)-(1.2) under the conditions (C1-2), (C2-2), (C3) or (C4-2). This finishes the proof of the theorem theorem 4.1  $\square$

## 5. BLOW UP IN FINITE TIME FOR THE SOLUTION OF THE NL SYSTEM.

We consider here the system (1.1)-(1.2) under the condition C5-1 with  $f_1 = 0$  and  $a = -b = -d_2 < 0$ . That is

$$(5.1) \quad \eta_t - b\eta_{xxt} = -u_x + bu_{xxx} \quad x, t \in \mathbb{R},$$

$$(5.2) \quad u_t - du_{xxt} + bu_{xxxxt} = -\eta_x + b\eta_{xxx} - (f_2(\eta, u))_x + b(f_2(\eta, u))_{xxx}$$

with  $\eta(x, 0) = \eta_0(x)$ ,  $u(x, 0) = u_0(x)$ ,  $\eta(\infty, t) = \eta_x(\infty, t) = \eta_{xx}(\infty, t) = u(\infty, t) = u_x(\infty, t) = u_{xx}(\infty, t) = 0$  and where,  $f_2(\eta, u) = u^{\alpha+1}$  or  $f_2(\eta, u) = \eta^{\alpha+1}$  or  $f_2(\eta, u) = u\eta + \frac{1}{2}\eta^2$ ,  $\alpha \geq 1$ .

Then, setting  $P_1 = 1 - b\frac{\partial^2}{\partial x^2}$  and  $P_2 = 1 - d\frac{\partial^2}{\partial x^2} + b\frac{\partial^4}{\partial x^4}$ , (5.1)-(5.2) becomes

$$(5.3) \quad \eta_t = -u_x \quad x, t \in \mathbb{R},$$

$$(5.4) \quad u_t = -P_2^{-1}P_1(\eta + f_2(\eta, u))_x.$$

We already know from the theorem 2.1 above the local existence of the solution of (5.3)-(5.4). We will prove now a blow up theorem of the solution to (5.3)-(5.4). Note that all the theorems in the sequel, work also, thanks to the symetry, for the solutions of (1.1)-(1.2) under the condions (C5-2).

Before giving the blow-up theorems, let us prove the following needed lemmas.

**Lemma 5.1.** *If there exist functions  $u_0(x) \in \mathbf{H}^{s+1}$ ,  $w_0(x) \in \mathbf{H}^{s+2}$ ,  $s > \frac{1}{2}$ , such that the initial values  $\eta(x, 0)$ ,  $\eta_t(x, 0)$ ,  $u(x, 0)$ , satisfy the relations*

$$\eta(x, 0) = (w_0(x))_x, \quad \eta_t(x, 0) = -(u_0(x))_x,$$

*then for all  $t \in [0, T]$ , the solution  $(\eta, u)$  of the Cauchy problem associated to (5.3)-(5.4) satisfies*

$$\eta(x, t) = (w(x, t))_x,$$

*with a corresponding evolution of  $w(x, t)$ ,  $\eta(x, t)$  satisfying the system*

$$(5.5) \quad w_t(x, t) = -u(x, t) \quad x, t \in \mathbb{R},$$

$$(5.6) \quad u_t = -P_2^{-1}P_1(\eta + f_2(\eta, u))_x.$$

*The couple of functions  $(w, u)$  belongs to  $C^1([0, T]; \mathbf{H}^{s+2}) \times C^1([0, T]; \mathbf{H}^{s+1})$*

*Proof.* From (5.3) we find

$$(5.7) \quad \eta(x, t) = \eta(x, 0) - \int_0^t (u(x, s))_x ds.$$

The term  $\eta(x, 0)$  is an  $x$ -derivative by hypothesis and  $\int_0^t u_x ds$  is an  $x$ -derivative. Therefore, there exists a  $w(x, t)$  such that  $\eta(x, t) = w_x$ . The second part of the lemma is proved by Theorem 2.1 above.  $\square$

We need also, for the result of blow-up, the following lemma proved in (Levine, 1974).

## GLOBAL EXISTENCE

**Lemma 5.2.** Suppose  $\psi(t)$ ,  $t \geq 0$ , is a positive, twice-differentiable function satisfying the inequality

$$\psi''\psi - (1 + \gamma)(\psi')^2 \geq 0$$

where  $\gamma > 0$ .

If  $\psi(0) > 0$  and  $\psi'(0) > 0$ , then  $\psi(t) \rightarrow \infty$  as  $t \rightarrow t_1 \leq \frac{\psi(0)}{\gamma\psi'(0)}$ ; ( $t_1$  is a positive constant).

Let us give now the blow-up theorem.

**Theorem 5.3.** (Blow-up for the solution of (5.3)-(5.4)).  
Suppose that there exists  $\gamma > 0$  such that

$$(5.8) \quad \eta(\eta + f_2(\eta, u)) \leq 2(1 + 2\gamma)F_2(\eta, u)$$

where  $F_2$  is such that

$$\frac{\partial F_2}{\partial \eta} = \eta + \lambda_0 f_2(u, \eta) + \sum_{j=1}^k \lambda_j u^j$$

$$\frac{\partial F_2}{\partial u} = \nu_0(\eta + f_2(\eta, u))$$

with  $\nu_0 \geq 0$ ,  $\lambda_j \geq 0$ ,  $\forall j, k \geq 0$  and where  $\lambda_0 = 1$  if  $f_2$  is not a purely function of  $u$ .

Suppose also that the initial values  $\eta(x, 0)$ ,  $\eta_t(x, 0)$ ,  $u(x, 0)$ , are chosen such that they satisfy

- $\eta(x, 0) = (w_0(x))_x$ ,  $\eta_t(x, 0) = -(u_0(x))_x$ , for some  $u_0(x) \in \mathbf{H}^{s+1}$ ,  $w_0(x) \in \mathbf{H}^{s+2}$ ,  $s > \frac{1}{2}$ ,
- $E(0) = \frac{1}{2}(u_0, P_1^{-1}P_2u_0) + \int_{\mathbf{R}} F_2(\eta_0, u_0)dx < 0$ .

Then, the solution  $(\eta, u)$  of (5.3)-(5.4) blows-up in finite time.

*Proof.* Proof of the theorem 5.3. For example: If  $f_2(\eta, u) = u^{\alpha+1}$  then  $F_2(\eta, u) = \frac{1}{2}\eta^2 + u\eta + \frac{1}{\alpha+2}u^{\alpha+2}$ , or if  $f_2(\eta, u) = \eta^{\alpha+1}$  then  $F_2(\eta, u) = \frac{1}{2}\eta^2 + \frac{1}{\alpha+2}\eta^{\alpha+2}$ , or if  $f_2(\eta, u) = \frac{1}{2}\eta^2 + u\eta$  then  $F_2(\eta, u) = \frac{1}{2}\eta^2 + u\eta + \frac{1}{2}\eta u^2 + \frac{1}{2}\eta^2 u + \frac{1}{6}\eta^3$ .

To begin the proof, let

$$E(t) = \frac{1}{2}(u, P_1^{-1}P_2u) + \int_{\mathbf{R}} F_2(\eta, u)dx.$$

We claim that

$$(5.9) \quad E(t) = E(0) = \frac{1}{2}(u_0, P_1^{-1}P_2u_0) + \int_{\mathbf{R}} F_2(\eta_0, u_0)dx < 0.$$

DÉ GODEFROY AKMEL

Indeed, thanks to (5.3), (5.4) and integrations by parts, we have

$$\begin{aligned}
 (5.10) \quad E'(t) &= (u_t, P_1^{-1} P_2 u) + \int_{\mathbf{R}} \left\{ \eta_t \frac{\partial F_2}{\partial \eta}(\eta, u) + u_t \frac{\partial F_2}{\partial u}(\eta, u) \right\} dx \\
 &= -((\eta + f_2(\eta, u))_x, u) - \int_{\mathbf{R}} u_x \frac{\partial F_2}{\partial \eta}(\eta, u) dx \\
 &\quad - \int_{\mathbf{R}} P_2^{-1} P_1 (\eta + f_2(\eta, u))_x \frac{\partial F_2}{\partial u}(\eta, u) dx \\
 &= -((\eta + f_2(\eta, u))_x, u) - \int_{\mathbf{R}} u_x (\eta + \lambda_0 f_2(u, \eta) + \sum_{j=1}^k \lambda_j u^j) dx \\
 &\quad - \nu_0 \int_{\mathbf{R}} P_2^{-1} P_1 (\eta + f_2(\eta, u))_x (\eta + f_2(\eta, u)) dx = 0.
 \end{aligned}$$

Now, define  $\psi(t)$  as follows.

$$\psi(t) = (w, P_1^{-1} P_2 w) + \beta_0 (t + t_0)^2$$

where  $\beta_0$  and  $t_0$  are non-negatives constants to be defined later. Suppose that the maximal time of existence is infinite. A contradiction will be obtained by lemma 5.2. The condition necessary to apply the lemma 5.2 is that  $\psi''\psi - (1 + \gamma)(\psi')^2 \geq 0$ . We have by hypotheses of theorem 5.3, and thanks to the lemma 5.1 and integration by parts,

$$\psi'(t) = 2(w_t, P_1^{-1} P_2 w) + 2\beta_0(t + t_0) = 2(-u, P_1^{-1} P_2 w) + 2\beta_0(t + t_0),$$

and

$$\begin{aligned}
 \psi''(t) &= 2(u, P_1^{-1} P_2 u) + 2(P_2^{-1} P_1 (\eta + f_2(\eta, u))_x, P_1^{-1} P_2 w) + 2\beta_0 \\
 &= 2(u, P_1^{-1} P_2 u) - 2(\eta + f_2(\eta, u), \eta) + 2\beta_0 \\
 &= 2(u, P_1^{-1} P_2 u) - 2 \int_{\mathbf{R}} [\eta(\eta + f_2(\eta, u)) - (2 + 4\gamma)F_2(\eta, u)] dx \\
 &\quad - 2(2 + 4\gamma)E(t) + (2 + 4\gamma)(u, P_1^{-1} P_2 u) + 2\beta_0 \\
 &\geq 4(1 + \gamma)(u, P_1^{-1} P_2 u) - 2(2 + 4\gamma)E(t) + 2\beta_0.
 \end{aligned}$$

## GLOBAL EXISTENCE

It follows that, since  $\psi(t) \geq 0$ , and using the Schwartz inequality and the Parseval formula,

$$\begin{aligned}
\psi''\psi - (1+\gamma)(\psi')^2 &\geq [4(1+\gamma)(u, P_1^{-1}P_2u) - 2(2+4\gamma)E(t) + 2\beta_0]\psi(t) \\
&\quad - 4(1+\gamma)[(-u, P_1^{-1}P_2w) + \beta_0(t+t_0)]^2 \\
&\geq [4(1+\gamma)(u, P_1^{-1}P_2u) - 2(2+4\gamma)E(t) + 2\beta_0]\psi(t) \\
&\quad - 4(1+\gamma)[(u, P_1^{-1}P_2u)(w, P_1^{-1}P_2w) \\
&\quad + 2(u, P_1^{-1}P_2u)^{\frac{1}{2}}(w, P_1^{-1}P_2w)^{\frac{1}{2}}\beta_0(t+t_0)] \\
&\quad - 4(1+\gamma)\beta_0^2(t+t_0)^2 \\
&\geq [4(1+\gamma)(u, P_1^{-1}P_2u) - 2(2+4\gamma)E(t) + 2\beta_0]\psi(t) \\
&\quad - 4(1+\gamma)[(u, P_1^{-1}P_2u)(w, P_1^{-1}P_2w) \\
&\quad + 2(u, P_1^{-1}P_2u)^{\frac{1}{2}}(w, P_1^{-1}P_2w)^{\frac{1}{2}}\beta_0(t+t_0)] \\
&\quad - 4(1+\gamma)\beta_0\psi(t) + 4(1+\gamma)\beta_0(w, P_1^{-1}P_2w) \\
&\geq -2(1+2\gamma)[2E(t) + \beta_0]\psi(t) + 4(1+\gamma)(u, P_1^{-1}P_2u)\beta_0(t+t_0)^2 \\
&\quad - 8(1+\gamma)(u, P_1^{-1}P_2u)^{\frac{1}{2}}(w, P_1^{-1}P_2w)^{\frac{1}{2}}\beta_0(t+t_0) \\
&\quad + 4(1+\gamma)\beta_0(w, P_1^{-1}P_2w) \\
&\geq -2(1+2\gamma)[2E(t) + \beta_0]\psi(t) \\
&\quad + 4(1+\gamma)(u, P_1^{-1}P_2u)\beta_0(t+t_0)^2 - 4(1+\gamma)(u, P_1^{-1}P_2u)\beta_0(t+t_0)^2 \\
&\quad - 4(1+\gamma)\beta_0(w, P_1^{-1}P_2w) + 4(1+\gamma)\beta_0(w, P_1^{-1}P_2w) \\
&\geq -2(1+2\gamma)[2E(t) + \beta_0]\psi(t).
\end{aligned}$$

Thereby, since  $E(t) = E(0) < 0$ , it follows by taking  $\beta_0 = -2E(0)$ , that  $\psi''\psi - (1+\gamma)(\psi')^2 \geq 0$ . Also  $\psi'(0) = 2(-u_0, P_1^{-1}P_2w_0) + 2\beta_0t_0 > 0$  if  $t_0$  is sufficiently large. Thus, by lemma 5.2,  $\psi(t)$  becomes infinite at a time  $T_1$  at most equal to  $t_{\beta_0} = \frac{\psi(0)}{\gamma\psi'(0)} < \infty$ . Therefore, we have a contradiction with the fact that the maximal time of existence is infinite. Henceforth, there is blow up in finite time, and the maximal time of existence is finite.  $\square$

## REFERENCES

1. (2007)M.Chen & O.Goubet, (2007), Long-time asymptotic behavior of dissipative Boussinesq systems, *JDiscrete and continuous dynamical systems - serie A*, **7**, Num 3.
2. (2004)J.L.Bona, M.Chen, J.C.Saut, (2004). Boussinesq equations and others systems for small-amplitude long waves in nonlinear dispersive media: II. The nonlinear theory, *Nonlinearity* **17**, 925-952.
3. (1974) H. Levin, (1974). Instability and non-existence of global solutions to non-linear wave equations of the form  $Pu_{tt} = -Au + F(u)$ , *Trans. Amer. Math. Soc.*, **92**, 1-21.

# Caputo Fractional Multivariate Opial type inequalities on spherical shells

George A. Anastassiou  
 Department of Mathematical Sciences  
 University of Memphis  
 Memphis, TN 38152 U.S.A.  
 ganastss@memphis.edu

## Abstract

Here is introduced the concept of Caputo fractional radial derivative for a function defined on a spherical shell. Using polar coordinates we are able to derive multivariate Opial type inequalities over a spherical shell of  $\mathbb{R}^N$ ,  $N \geq 2$ , by studying the topic in all possibilities. Our results involve one, two or more functions. We produce many univariate Caputo fractional Opial type inequalities several of these used to establish results on the shell. We give application to prove uniqueness of solution of a general partial differential equation on the shell. Also we apply our results for Riemann-Liouville fractional derivatives.

**2000 Mathematics Subject Classification:** 26A33, 26D10, 26D15.

**Key Words and Phrases:** Opial inequality, fractional inequality, Caputo fractional derivative, radial derivative, univariate and multivariate inequality .

## 1 Introduction

This work is inspired by articles of Opial [22], Bessack [13], and Anastassiou-Koliha-Pecaric [11], [12], and Anastassiou [9], [10]. We would like to mention

**Theorem A.** (Opial [22], 1960) *Let  $c > 0$  and  $y(x)$  be real, continuously differentiable on  $[0, c]$ , with  $y(0) = y(c) = 0$ . Then*

$$\int_0^c |y(x) y'(x)| dx \leq \frac{c}{4} \int_0^c (y'(x))^2 dx.$$

*Equality holds for the function*

$$y(x) = x \text{ on } [0, c/2]$$

and

$$y(x) = c - x \text{ on } [c/2, c].$$

The next result implies Theorem A and is very useful to applications. Also it is our main motivation.

**Theorem B.** (Bessack [13], 1962) *Let  $b > 0$ . If  $y(x)$  is real, continuously differentiable on  $[0, b]$ , and  $y(0) = 0$ , then*

$$\int_0^b |y(x) y'(x)| dx \leq \frac{b}{2} \int_0^b (y'(x))^2 dx.$$

*Equality holds only for  $y = mx$ , where  $m$  is a constant.*

Opial type inequalities usually find applications in establishing uniqueness of solution of initial value problems for differential equations and their systems, see Willett [27]. In this article we present a series of various Caputo fractional multivariate Opial type inequalities over spherical shells. To achieve our goal we use polar coordinates, and we introduce and use the Caputo fractional radial derivative. We work on the spherical shell, and not on the ball, because a radial derivative can not be defined at zero. So, we reduce the problem to a univariate one.

Therefore we derive and use a large array of univariate Opial type inequalities involving Caputo fractional derivatives; these are Caputo fractional derivatives defined at arbitrary anchor point  $a \in \mathbb{R}$ . In our results we involve one, two, or several functions. But first we need to develop an extensive background in two parts, then follow the main results.

At the end we give application proving uniqueness of solution for a general PDE initial value problem. Also we reestablish our results by involving Riemann-Liouville fractional derivatives defined at an arbitrary anchor point.

In this article to build our background regarding Caputo fractional derivative we use the excellent monograph [17].

Caputo derivative was introduced in 1967, see [14], also see [15], [16].

It happens that the Riemann-Liouville fractional derivative has some disadvantages when is to model real-world phenomena with fractional differential equations. One reason is that the initial conditions there involve fractional derivatives that are difficult to connect with actual data, etc. However Caputo fractional derivative modelling involves initial conditions that are described by ordinary derivatives, much easier to write out of real world data. So more and more in recent years the Caputo version is usually preferred when physical models are described, because the physical interpretation of the prescribed data is clear, and therefore it is in general easier possible to gather these data, e.g. by appropriate measurements. Also from the pure mathematics side there are reasons to prefer recently more and more the Caputo fractional derivative.

## 2 Background-I

Here we follow [17].

We start with

**Definition 1.** Let  $\nu \geq 0$ , the operator  $J_a^\nu$ , defined on  $L_1[a, b]$  by

$$J_a^\nu f(x) := \frac{1}{\Gamma(\nu)} \int_a^x (x-t)^{\nu-1} f(t) dt \quad (1)$$

for  $a \leq x \leq b$ , is called the Riemann-Liouville fractional integral operator of order  $\nu$ .

For  $\nu = 0$ , we set  $J_a^0 := I$ , the identity operator. Here  $\Gamma$  stands for the gamma function.

**Theorem 2.** ([17]) Let  $f \in L_1[a, b]$ ,  $\nu > 0$ . Then, the integral  $J_a^\nu f(x)$  exists for almost every  $x \in [a, b]$ .

Moreover,  $J_a^\nu f \in L_1([a, b])$ .

We need

**Theorem 3.** ([17]) Let  $m, n \geq 0$ ,  $\Phi \in L_1([a, b])$ .

Then

$$J_a^m J_a^n \Phi = J_a^{m+n} \Phi \quad (2)$$

holds almost everywhere on  $[a, b]$ . If additionally  $\Phi \in C([a, b])$  or  $m+n \geq 1$ , then the identity holds everywhere on  $[a, b]$ .

We give

**Definition 4.** ([17]) Let  $\nu \in \mathbb{R}_+$  and  $m = \lceil \nu \rceil$ ,  $\lceil \cdot \rceil$  is the ceiling of number. The operator  $D_a^\nu$ , defined by

$$D_a^\nu f := D^m J_a^{m-\nu} f, \quad D := \frac{d}{dx}, \quad (3)$$

is called the Riemann-Liouville fractional differential operator of order  $\nu$ .

For  $\nu = 0$ , we set  $D_a^0 := I$ , the identity operator.

If  $\nu \in \mathbb{N}$  then  $D_a^\nu f = f^{(\nu)}$ , the ordinary  $\nu$  order derivative.

Next we give

**Definition 5.** (p.37, [17]) Let  $\nu \geq 0$  and  $n := \lceil \nu \rceil$ ,  $a \in \mathbb{R}$ . Then, we define the operator

$$\hat{D}_a^\nu f := J_a^{n-\nu} f^{(n)}, \quad (4)$$

whenever  $f^{(n)} \in L_1([a, b])$ .

Also we need

**Theorem 6.** (p.37, [17]) Let  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$ . Moreover assume that  $f \in AC^n([a, b])$  (space of functions with absolutely continuous  $(n-1)$ -st derivative). Then

$$\hat{D}_a^\nu f = D_a^\nu (f - T_{n-1}(f; a)), \quad \text{a.e. on } [a, b], \quad (5)$$

where

$$T_{n-1}(f; a)(x) := \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (x-a)^k, \quad x \in [a, b], \quad (6)$$

is the Taylor polynomial of degree  $n - 1$  of  $f$ , centered at  $a$ .

Next we give the definition of Caputo fractional derivative ([17]).

**Definition 7.** (p.38, [17]) Assume that  $f$  is such that  $D_a^\nu (f - T_{n-1}(f; a))(x)$  exists for some  $x \in [a, b]$ . Then we define the Caputo fractional derivative by

$$D_{*a}^\nu f(x) := D_a^\nu (f - T_{n-1}(f; a))(x). \quad (7)$$

So the above definition applies to all points  $x \in [a, b] : D_a^\nu (f - T_{n-1}(f; a))(x) \in \mathbb{R}$ .

We have

**Corollary 8.** Let  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f \in AC^n([a, b])$ . Then the Caputo fractional derivative

$$D_{*a}^\nu f(x) = \frac{1}{\Gamma(n - \nu)} \int_a^x (x - t)^{n - \nu - 1} f^{(n)}(t) dt \quad (8)$$

exists almost everywhere for  $x$  in  $[a, b]$ .

We have

**Corollary 9.** Let  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f \in AC^n([a, b])$ . Then,  $D_{*a}^\nu f$  exists iff  $D_a^\nu f$  exists.

**Proof.** By linearity of  $D_a^\nu$  operator and assumption.  $\square$

We need

**Lemma 10.** ([17]) Let  $\nu \geq 0$ ,  $n = \lceil \nu \rceil$ . Assume that  $f$  is such that both  $D_{*a}^\nu f$  and  $D_a^\nu f$  exist.

Then,

$$D_{*a}^\nu f(x) = D_a^\nu f(x) - \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{\Gamma(k - \nu + 1)} (x - a)^{k - \nu}. \quad (9)$$

**Lemma 11.** ([17]) All as in Lemma 10.

Additionally assume that  $f^{(k)}(a) = 0$  for  $k = 0, 1, \dots, n - 1$ . Then,

$$D_{*a}^\nu f = D_a^\nu f. \quad (10)$$

In conclusion

**Corollary 12.** Let  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f \in AC^n([a, b])$ ,  $D_{*a}^\nu f$  exists or  $D_a^\nu f$  exists, and  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n - 1$ . Then

$$D_a^\nu f = D_{*a}^\nu f. \quad (11)$$

We need the following Taylor-Caputo formula

**Theorem 13.** (p.40, [17]) Let  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f \in AC^n([a, b])$ . Then

$$f(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (x - a)^k + J_a^\nu D_{*a}^\nu f(x), \quad (12)$$

$\forall x \in [a, b]$ .



Clearly here  $J_a^\nu D_{*a}^\nu f \in AC^n([a, b])$ .

**Corollary 14.** Let  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f \in AC^n([a, b])$ , and  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ . Then

$$f(x) = J_a^\nu D_{*a}^\nu f(x) = \frac{1}{\Gamma(\nu)} \int_a^x (x-t)^{\nu-1} D_{*a}^\nu f(t) dt. \quad (13)$$

We need

**Lemma 15.** Let  $\nu \geq \gamma+1$ ,  $\gamma \geq 0$ . Call  $n := \lceil \nu \rceil$ ,  $m := \lceil \gamma \rceil$ . Then  $n-m \geq 1$ , i.e.  $m \leq n-1$ .

**Proof.** Clearly  $\nu \geq 1$  and  $\nu > \gamma$ ,  $\nu - \gamma \geq 1$ . By  $\gamma+1 > m$  we get  $\nu > m$ , and  $n > m$ , that is  $\nu - m > 0$  and  $n - m > 0$ .

We see that  $\nu \geq \gamma+1 \geq \lceil \gamma \rceil + 1$ , where  $\lceil \cdot \rceil$  is the integral part. Thus  $\nu \geq (\lceil \gamma \rceil + 1) \in \mathbb{N}$  and  $\nu \geq \lceil \nu \rceil \geq \lceil \gamma \rceil + 1$ .

Therefore

$$\lceil \nu \rceil - \lceil \gamma \rceil \geq 1, \quad (14)$$

to be used next.

We distinguish the following cases.

i) Let  $\nu, \gamma \notin \mathbb{N}$ , then  $\lceil \nu \rceil = \lceil \nu \rceil + 1$ ,  $\lceil \gamma \rceil = \lceil \gamma \rceil + 1$ . By (14) we get  $(\lceil \nu \rceil + 1) - (\lceil \gamma \rceil + 1) \geq 1$ . Hence  $n - m \geq 1$ .

ii) Let  $\nu, \gamma \in \mathbb{N}$ , then  $\lceil \nu \rceil = \nu$ ,  $\lceil \gamma \rceil = \gamma$ . So by (14)  $n - m \geq 1$ .

iii) Let  $\nu \notin \mathbb{N}$ ,  $\gamma \in \mathbb{N}$ . Then  $n = \lceil \nu \rceil = \lceil \nu \rceil + 1$ ,  $m = \lceil \gamma \rceil = \gamma$ . Hence by (14) we have  $(\lceil \nu \rceil + 1) - m \geq 1$ , and  $\lceil \nu \rceil - m \geq 2 > 1$ . Hence  $n - m > 1$ .

iv) Let  $\nu \in \mathbb{N}$ ,  $\gamma \notin \mathbb{N}$ . Then  $1 + \gamma < \lceil \gamma \rceil + 1 = \lceil \gamma + 1 \rceil$ , and  $1 + \gamma \leq \nu \in \mathbb{N}$  by assumption.

Therefore  $\lceil \gamma \rceil + 1 \leq \nu$ , and  $\nu - \lceil \gamma \rceil \geq 1$ . So that again  $n - m \geq 1$ .

Claim is proved in all cases.  $\square$

We present the representation theorem.

**Theorem 16.** Let  $\nu \geq \gamma+1$ ,  $\gamma \geq 0$ . Call  $n := \lceil \nu \rceil$ ,  $m := \lceil \gamma \rceil$ . Assume  $f \in AC^n([a, b])$ , such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f \in L_\infty(a, b)$ .

Then

$$D_{*a}^\gamma f \in C([a, b]), \quad D_{*a}^\gamma f(x) = J_a^{m-\gamma} f^{(m)}(x), \quad (15)$$

and

$$D_{*a}^\gamma f(x) = \frac{1}{\Gamma(\nu-\gamma)} \int_a^x (x-t)^{\nu-\gamma-1} D_{*a}^\nu f(t) dt, \quad (16)$$

$\forall x \in [a, b]$ .

**Proof.** If  $\gamma = 0$  then (16) collapses to (13), also (15) is clear. So we assume  $\gamma > 0$ . By Lemma 15 we have  $m \leq n-1$ . By the assumption  $f \in AC^n([a, b])$  we get that  $f \in C^{n-1}([a, b])$  and thus  $f \in C^m([a, b])$ .

By Lemma 3.7, p.41 of [17] we get that  $D_{*a}^\gamma f = J_a^{m-\gamma} f^{(m)} \in C([a, b])$  and  $D_{*a}^\gamma f(a) = 0$ , for  $\gamma \notin \mathbb{N}$ . Clearly the last statement is true also when  $\gamma \in \mathbb{N}$ , so proving (15) and first claim.

To remind, we have  $\nu - m > 0$ , and  $\nu - 1 > 0$  by  $\gamma > 0$ . Using  $\Gamma(p+1) = p\Gamma(p)$ ,  $p > 0$ , (13) and Theorem 7 of [9], we obtain

$$f^{(m)}(x) = J_a^{\nu-m} D_{*a}^\nu f(x), \quad \forall x \in [a, b]. \quad (17)$$

Therefore we get

$$\begin{aligned} D_{*a}^\gamma f(x) &= J_a^{m-\gamma} f^{(m)}(x) \stackrel{(17)}{=} J_a^{m-\gamma} J_a^{\nu-m} D_{*a}^\nu f(x) \\ &\quad (\text{by Theorem 2.2, p.14 of [17], and } \nu - \gamma \geq 1) \\ &= J_a^{\nu-\gamma} D_{*a}^\nu f(x), \quad \forall x \in [a, b]. \end{aligned}$$

That is proving (16).  $\square$

We also give the representation theorem.

**Theorem 17.** *Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ . Call  $n := \lceil \nu \rceil$ ,  $m := \lceil \gamma \rceil$ . Let  $f \in AC^n([a, b])$ , such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ . Assume there exists  $D_a^\nu f(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$ , and  $D_a^\nu f \in L_\infty(a, b)$ . Then*

$$D_a^\gamma f \in C([a, b]), \quad D_a^\gamma f(x) = J_a^{m-\gamma} f^{(m)}(x), \quad (18)$$

$\forall x \in [a, b]$ ,

$$D_a^\gamma f(x) = \frac{1}{\Gamma(\nu-\gamma)} \int_a^x (x-t)^{\nu-\gamma-1} D_a^\nu f(t) dt, \quad (19)$$

$\forall x \in [a, b]$ .

**Proof.** By Corollaries 9 and 12 we get existing  $D_{*a}^\nu f(x) \in \mathbb{R}$ , and that  $D_{*a}^\nu f(x) = D_a^\nu f(x)$ ,  $\forall x \in [a, b]$ . That is  $D_{*a}^\nu f \in L_\infty(a, b)$  and by (16) we have

$$D_{*a}^\gamma f(x) = \frac{1}{\Gamma(\nu-\gamma)} \int_a^x (x-t)^{\nu-\gamma-1} D_a^\nu f(t) dt, \quad \forall x \in [a, b]. \quad (20)$$

Since  $D_{*a}^\gamma f \in C([a, b])$  we get  $D_{*a}^\gamma f(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$ . And since  $f \in C^m([a, b])$  then  $f \in AC^m([a, b])$ . By Corollary 9  $D_a^\gamma f$  exists. Also  $f^{(k)}(a) = 0$ , for  $k = 0, 1, \dots, m-1$ . Thus by Corollary 12 we obtain  $D_a^\gamma f(x) = D_{*a}^\gamma f(x)$ ,  $\forall x \in [a, b]$ . Now by (20) we have established (19).  $\square$

## 3 Main Results

### 3.1 Results involving one function

We present

**Theorem 18.** *Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ . Call  $n := \lceil \nu \rceil$  and assume  $f \in AC^n([a, b])$  such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f \in L_\infty(a, b)$ .*

*Let  $p, q > 1$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ ,  $a \leq x \leq b$ .*

Then

$$\begin{aligned} & \int_a^x |D_{*a}^\gamma f(\omega)| |(D_{*a}^\nu f)(\omega)| d\omega \leq \\ & \frac{(x-a)^{\frac{p\nu-p\gamma-p+2}{p}}}{(\sqrt[p]{2}) \Gamma(\nu-\gamma) ((p\nu-p\gamma-p+1)(p\nu-p\gamma-p+2))^{1/p}} \\ & \cdot \left( \int_a^x |D_{*a}^\nu f(\omega)|^q d\omega \right)^{2/q}. \end{aligned} \quad (21)$$

**Proof.** Similar to Theorem 25.2, p.545, [2], and Theorem 2.1 of [11].  $\square$

A related extreme case comes next.

**Proposition 19.** All as in Theorem 18, but with  $p = 1$  and  $q = \infty$ , we find

$$\begin{aligned} & \int_a^x |D_{*a}^\gamma f(\omega)| |D_{*a}^\nu f(\omega)| d\omega \leq \\ & \frac{(x-a)^{\nu-\gamma+1}}{\Gamma(\nu-\gamma+2)} \left( \|D_{*a}^\nu f\|_{\infty, (a, x)} \right)^2. \end{aligned} \quad (22)$$

**Proof.** Similar to Proposition 25.1, p.547, [2].  $\square$

The converse of (21) follows.

**Theorem 20.** Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ . Call  $n := \lceil \nu \rceil$  and assume  $f \in AC^n([a, b])$  such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f, \frac{1}{D_{*a}^\nu f} \in L_\infty(a, b)$ . Suppose that  $D_{*a}^\nu f$  is of fixed sign a.e. in  $[a, b]$ . Let  $p, q$  such that  $0 < p < 1$ ,  $q < 0$  and  $\frac{1}{p} + \frac{1}{q} = 1$ ,  $a \leq x \leq b$ . Then

$$\begin{aligned} & \int_a^x |D_{*a}^\gamma f(\omega)| |D_{*a}^\nu f(\omega)| d\omega \geq \\ & \frac{(x-a)^{\frac{p\nu-p\gamma-p+2}{p}}}{(\sqrt[p]{2}) \Gamma(\nu-\gamma) ((p\nu-p\gamma-p+1)(p\nu-p\gamma-p+2))^{1/p}} \\ & \cdot \left( \int_a^x |D_{*a}^\nu f(\omega)|^q d\omega \right)^{2/q}. \end{aligned} \quad (23)$$

**Proof.** Similar to Theorem 25.3, p.547, [2], and Theorem 2.3 of [11].  $\square$

We give

**Theorem 21.** Let  $\nu \geq 2$ ,  $k \geq 0$ ,  $\nu \geq k + 2$ . Call  $n := \lceil \nu \rceil$  and assume  $f \in AC^n([a, b])$  such that  $f^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f \in L_\infty(a, b)$ . Let  $p, q > 1$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ ,  $a \leq x \leq b$ . Then

$$\int_a^x |D_{*a}^k f(\omega)| |D_{*a}^{k+1} f(\omega)| d\omega \leq$$

$$\frac{(x-a)^{\frac{2(p\nu-pk-p+1)}{p}}}{2(\Gamma(\nu-k))^2(p\nu-pk-p+1)^{2/p}} \cdot \left( \int_a^x |D_{*a}^\nu f(\omega)|^q d\omega \right)^{2/q}. \quad (24)$$

**Proof.** Similar to Theorem 25.4, p.549, [2], and Theorem 2.4 of [11].  $\square$

The extreme case follows.

**Proposition 22.** Under the assumptions of Theorem 21 when  $p = 1, q = \infty$  we find

$$\int_a^x |D_{*a}^k f(\omega)| |D_{*a}^{k+1} f(\omega)| d\omega \leq \frac{(x-a)^{2(\nu-k)} \left( \|D_{*a}^\nu f\|_{\infty, (a,x)} \right)^2}{2(\Gamma(\nu-k+1))^2}. \quad (25)$$

**Proof.** Similar to Proposition 25.2, p.551, [2].  $\square$

We give the related converse result.

**Theorem 23.** Let  $\nu \geq 2, k \geq 0, \nu \geq k+2$ . Call  $n := \lceil \nu \rceil$ . Assume  $f \in AC^n([a, b])$  such that  $f^{(j)}(a) = 0, j = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f, \frac{1}{D_{*a}^\nu f} \in L_\infty(a, b)$ . Suppose that  $D_{*a}^\nu f$  is of fixed sign a.e. in  $[a, b]$ . Let  $p, q$  such that  $0 < p < 1, q < 0$  and  $\frac{1}{p} + \frac{1}{q} = 1, a \leq x \leq b$ . Then

$$\int_a^x |D_{*a}^k f(\omega)| |D_{*a}^{k+1} f(\omega)| d\omega \geq \frac{(x-a)^{\frac{2(p\nu-pk-p+1)}{p}}}{2(\Gamma(\nu-k))^2(p\nu-pk-p+1)^{2/p}} \cdot \left( \int_a^x |D_{*a}^\nu f(\omega)|^q d\omega \right)^{2/q}. \quad (26)$$

**Proof.** Similar to Theorem 25.5, p.553 of [2].  $\square$

Next we present

**Theorem 24.** Let  $\gamma_i \geq 0, \nu \geq 1, \nu - \gamma_i \geq 1; i = 1, \dots, l, n := \lceil \nu \rceil$ , and  $f \in AC^n([a, b])$  such that  $f^{(k)}(a) = 0, k = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f \in L_\infty(a, b)$ . Here  $a \leq x \leq b; q_1(x), q_2(x)$  continuous functions on  $[a, b]$  such that  $q_1(x) \geq 0, q_2(x) > 0$  on  $[a, b]$ , and  $r_i > 0 : \sum_{i=1}^l r_i = r$ . Let  $s_1, s'_1 > 1 : \frac{1}{s_1} + \frac{1}{s'_1} = 1$  and  $s_2, s'_2 > 1 : \frac{1}{s_2} + \frac{1}{s'_2} = 1$ , and  $p > s_2$ .

Denote by

$$Q_1(x) := \left( \int_a^x (q_1(\omega))^{s'_1} d\omega \right)^{1/s'_1} \quad (27)$$

and

$$Q_2(x) := \left( \int_a^x (q_2(\omega))^{-s'_2/p} d\omega \right)^{r/s'_2}, \quad (28)$$

$$\sigma := \frac{p - s_2}{ps_2}. \quad (29)$$

Then

$$\begin{aligned} & \int_a^x q_1(\omega) \prod_{i=1}^l |D_{*a}^{\gamma_i} f(\omega)|^{r_i} d\omega \leq Q_1(x) Q_2(x) \\ & \prod_{i=1}^l \left\{ \frac{\sigma^{r_i \sigma}}{(\Gamma(\nu - \gamma_i))^{r_i} (\nu - \gamma_i - 1 + \sigma)^{r_i \sigma}} \right\} \\ & \cdot \frac{(x - a)^{\left( \sum_{i=1}^l (\nu - \gamma_i - 1) r_i + \sigma r + \frac{1}{s_1} \right)}}{\left( \left( \sum_{i=1}^l (\nu - \gamma_i - 1) r_i s_1 \right) + r s_1 \sigma + 1 \right)^{1/s_1}} \\ & \cdot \left( \int_a^x q_2(\omega) |D_{*a}^{\nu} f(\omega)|^p d\omega \right)^{r/p}. \end{aligned} \quad (30)$$

**Proof.** Similar to Theorem 26.1, p.567, [2], and Theorem 2.1 of [12].  $\square$

The counterpart of last theorem follows.

**Theorem 25.** Let  $\gamma_i \geq 0$ ,  $\nu \geq 1$ ,  $\nu - \gamma_i \geq 1$ ;  $i = 1, \dots, l$ ,  $n := \lceil \nu \rceil$ , and  $f \in AC^n([a, b])$  such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n - 1$ , and  $D_{*a}^{\nu} f, \frac{1}{D_{*a}^{\nu} f} \in L_{\infty}(a, b)$ . Here  $a \leq x \leq b$ ;  $q_1(x), q_2(x) > 0$  continuous functions on  $[a, b]$  and  $r_i > 0$ :  $\sum_{i=1}^l r_i = r$ . Let  $0 < s_1, s_2 < 1$  and  $s'_1, s'_2 < 0$  such that  $\frac{1}{s_1} + \frac{1}{s'_1} = 1$ ,  $\frac{1}{s_2} + \frac{1}{s'_2} = 1$ . Assume that  $D_{*a}^{\nu} f(t)$  is of fixed sign a.e. in  $[a, b]$ . Denote

$$Q_1(x) := \left( \int_a^x (q_1(\omega))^{s'_1} d\omega \right)^{1/s'_1}, \quad (31)$$

$$Q_2(x) := \left( \int_a^x (q_2(\omega))^{-s'_2} d\omega \right)^{r/s'_2}. \quad (32)$$

Set

$$\lambda := \frac{s_1 s_2}{s_1 s_2 - 1}. \quad (33)$$

Then

$$\int_a^x q_1(\omega) \left( \prod_{i=1}^l |D_{*a}^{\gamma_i} f(\omega)|^{r_i} \right) d\omega \geq$$

$$\begin{aligned}
& \frac{Q_1(x) Q_2(x)}{\prod_{i=1}^l \left\{ (\Gamma(\nu - \gamma_i))^{r_i} ((\nu - \gamma_i - 1) s_2^2 s_1 + 1) \left( \frac{r_i}{s_2^2 s_1} \right) \right\}} \\
& \cdot \frac{(x-a) \{ (\sum_{i=1}^l r_i ((\nu - \gamma_i - 1) s_1 + s_2^{-2})) + 1 \}^{1/s_1}}{\left\{ \left( \sum_{i=1}^l r_i ((\nu - \gamma_i - 1) s_1 + s_2^{-2}) \right) + 1 \right\}^{1/s_1}} \\
& \cdot \left( \int_a^x q_2^{\lambda s_2}(\omega) |D_{*a}^\nu f(\omega)|^{\lambda s_2} d\omega \right)^{r/\lambda s_2}. \tag{34}
\end{aligned}$$

**Proof.** Similar to Theorem 26.2, p.570, [2], and Theorem 2.3 of [12].  $\square$

A related extreme case comes next for  $p = 1$  and  $q = \infty$ .

**Theorem 26.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ;  $i = 1, \dots, l$ . Call  $n := \lceil \nu \rceil$  and assume  $f \in AC^n([a, b])$  such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f \in L_\infty(a, b)$ . Here  $a \leq x \leq b$ , with  $0 \leq \tilde{q}(\omega) \in L_\infty(a, b)$  and  $r_i > 0$ :  $\sum_{i=1}^l r_i = r$ . Then

$$\begin{aligned}
& \int_a^x \tilde{q}(\omega) \prod_{i=1}^l |D_{*a}^{\gamma_i} f(\omega)|^{r_i} d\omega \leq \\
& \left\{ \frac{\|\tilde{q}\|_{\infty, (a, x)} \left( \|D_{*a}^\nu f\|_{\infty, (a, x)} \right)^r}{\prod_{i=1}^l (\Gamma(\nu - \gamma_i + 1))^{r_i}} \right\} \cdot \\
& \left\{ \frac{(x-a)^{r\nu - \sum_{i=1}^l r_i \gamma_i + 1}}{(r\nu - \sum_{i=1}^l r_i \gamma_i + 1)} \right\}. \tag{35}
\end{aligned}$$

**Proof.** Similar to Proposition 26.1 of [2], p.573 and Theorem 2.2 of [12].

$\square$

We continue with the interesting

**Theorem 27.** Let  $k \geq 0$ ,  $\gamma \geq 1$ ,  $\nu \geq 2$ ,  $n := \lceil \nu \rceil$ , such that  $\nu - \gamma \geq 1$ ,  $\gamma - k \geq 1$ , and  $f \in AC^n([a, b])$  such that  $f^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f \in L_\infty(a, b)$ . Here  $a \leq x \leq b$ ,  $p, q > 1$ :  $\frac{1}{p} + \frac{1}{q} = 1$ . Then

$$\begin{aligned}
& \int_a^x |D_{*a}^\gamma f(\omega)| |D_{*a}^k f(\omega)| d\omega \leq \\
& \frac{2^{-1/p} (x-a)^{(2\nu - k - \gamma - 1 + \frac{2}{q})}}{\Gamma(\nu - k) \Gamma(\nu - \gamma + 1) ((\nu - \gamma)q + 1)^{1/q}} \\
& \cdot \frac{\left( \int_a^x |D_{*a}^\nu f(\omega)|^p d\omega \right)^{2/p}}{(2\nu q - kq - \gamma q - q + 2)^{1/q}}. \tag{36}
\end{aligned}$$

**Proof.** Similar to Theorem 26.3, p.574, [2], and Theorem 2.5 of [12].  $\square$

We give

**Theorem 28.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, \dots, k \in \mathbb{N} - \{1\}$ ,  $n := \lceil \nu \rceil$ . Assume  $f \in AC^n([a, b])$  such that  $f^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ , and  $D_{*a}^\nu f \in L_\infty(a, b)$ . Here  $a \leq x \leq b$ ,  $\gamma := \sum_{i=1}^k \gamma_i$ . Let  $p, q > 1$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ . Furthermore, suppose that  $|D_{*a}^\nu f(t)|$  is decreasing on  $[a, x]$ . Then

$$\begin{aligned} & \int_a^x \prod_{i=1}^k |D_{*a}^{\gamma_i} f(\omega)| d\omega \leq \\ & \frac{p(x-a)^{\left(\frac{1+k\nu p - \gamma p}{p}\right)}}{\prod_{i=1}^k \Gamma(\nu - \gamma_i) (k\nu p - \gamma p - kp + 1)^{1/p}} \\ & \cdot \frac{\left(\int_a^x |D_{*a}^\nu f(t)|^{kq} dt\right)^{1/q}}{(k\nu p - \gamma p - kp + p + 1)}. \end{aligned} \quad (37)$$

**Proof.** Similar to Theorem 26.6, p.581 of [2], and Theorem 2.6 of [12].  $\square$

The extreme case follows

**Theorem 29.** All as in Theorem 28, but  $p = 1$ ,  $q = \infty$ . Then

$$\begin{aligned} & \int_a^x \prod_{i=1}^k |D_{*a}^{\gamma_i} f(\omega)| d\omega \leq \\ & \frac{(x-a)^{k\nu - \gamma + 1} \left(\|D_{*a}^\nu f\|_{\infty, (a, x)}\right)^k}{\left(\prod_{i=1}^k \Gamma(\nu - \gamma_i)\right) (k\nu - \gamma - k + 1) (k\nu - \gamma + 1)}. \end{aligned} \quad (38)$$

**Proof.** Similar to Proposition 26.4, p.582 of [2], and Theorem 2.7 of [12].

$\square$

### 3.2 Results involving two functions

We present

**Theorem 30.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, 2$ ,  $n := \lceil \nu \rceil$ , and  $f_1, f_2 \in AC^n([a, b])$  such that  $f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ . Consider also  $p(t) > 0$  and  $q(t) \geq 0$ , with all  $p(t)$ ,  $\frac{1}{p(t)}$ ,  $q(t) \in L_\infty(a, b)$ . Further assume  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ .

Let  $\lambda_\nu > 0$  and  $\lambda_\alpha, \lambda_\beta \geq 0$  such that  $\lambda_\nu < p$ , where  $p > 1$ . Set

$$P_k(\omega) := \int_a^\omega (\omega - t)^{\frac{(\nu - \gamma_k - 1)p}{p-1}} (p(t))^{-\frac{1}{(p-1)}} dt, \quad (39)$$

$k = 1, 2, a \leq x \leq b;$

$$A(\omega) := \frac{q(\omega) (P_1(\omega))^{\lambda_\alpha \left(\frac{p-1}{p}\right)} (P_2(\omega))^{\lambda_\beta \left(\frac{p-1}{p}\right)} (p(\omega))^{-\lambda_\nu/p}}{(\Gamma(\nu - \gamma_1))^{\lambda_\alpha} (\Gamma(\nu - \gamma_2))^{\lambda_\beta}}, \quad (40)$$

$$A_0(x) := \left( \int_a^x (A(\omega))^{\frac{p}{p-\lambda_\nu}} d\omega \right)^{\frac{p-\lambda_\nu}{p}}, \quad (41)$$

and

$$\delta_1 := \begin{cases} 2^{1-\left(\frac{\lambda_\alpha+\lambda_\nu}{p}\right)}, & \text{if } \lambda_\alpha + \lambda_\nu \leq p, \\ 1, & \text{if } \lambda_\alpha + \lambda_\nu \geq p. \end{cases} \quad (42)$$

If  $\lambda_\beta = 0$ , we obtain that

$$\begin{aligned} & \int_a^x q(\omega) \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \quad \left. |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \leq \\ & \quad (A_0(x)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{\lambda_\alpha + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_1 \\ & \quad \left[ \int_a^x p(\omega) [|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p] d\omega \right]^{\left(\frac{\lambda_\alpha+\lambda_\nu}{p}\right)}. \end{aligned} \quad (43)$$

**Proof.** Similar to Theorem 2 of [3] and Theorem 4 of [9].  $\square$

It follows the counterpart of the last theorem.

**Theorem 31.** All here as in Theorem 30.

Denote

$$\delta_3 := \begin{cases} 2^{\lambda_\beta/\lambda_\nu} - 1, & \text{if } \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\beta \leq \lambda_\nu. \end{cases} \quad (44)$$

If  $\lambda_\alpha = 0$ , then

$$\begin{aligned} & \int_a^x q(\omega) \left[ |D_{*a}^{\gamma_2} f_2(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \quad \left. |D_{*a}^{\gamma_2} f_1(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \leq \\ & \quad (A_0(x)|_{\lambda_\alpha=0}) 2^{\frac{p-\lambda_\nu}{p}} \left( \frac{\lambda_\nu}{\lambda_\beta + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_3^{\lambda_\nu/p} \\ & \quad \left( \int_a^x p(\omega) [|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p] d\omega \right)^{\left(\frac{\lambda_\nu+\lambda_\beta}{p}\right)}, \end{aligned} \quad (45)$$

all  $a \leq x \leq b$ .



**Proof.** Similar to Theorem 3 of [3] and Theorem 5 of [9].  $\square$

The complete case  $\lambda_\alpha, \lambda_\beta \neq 0$  follows.

**Theorem 32.** *All here as in Theorem 30.*

Denote

$$\tilde{\gamma}_1 := \begin{cases} 2^{\left(\frac{\lambda_\alpha + \lambda_\beta}{\lambda_\nu}\right)} - 1, & \text{if } \lambda_\alpha + \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\alpha + \lambda_\beta \leq \lambda_\nu, \end{cases} \quad (46)$$

and

$$\tilde{\gamma}_2 := \begin{cases} 1, & \text{if } \lambda_\alpha + \lambda_\beta + \lambda_\nu \geq p, \\ 2^{1 - \left(\frac{\lambda_\alpha + \lambda_\beta + \lambda_\nu}{p}\right)}, & \text{if } \lambda_\alpha + \lambda_\beta + \lambda_\nu \leq p. \end{cases} \quad (47)$$

Then

$$\begin{aligned} & \int_a^x q(\omega) \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_2} f_2(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \quad \left. |D_{*a}^{\gamma_2} f_1(\omega)|^{\lambda_\beta} |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \leq \\ & A_0(x) \left( \frac{\lambda_\nu}{(\lambda_\alpha + \lambda_\beta)(\lambda_\alpha + \lambda_\beta + \lambda_\nu)} \right)^{\lambda_\nu/p} \left[ \lambda_\alpha^{\lambda_\nu/p} \tilde{\gamma}_2 + 2^{\frac{p-\lambda_\nu}{p}} (\tilde{\gamma}_1 \lambda_\beta)^{\lambda_\nu/p} \right] \\ & \quad \cdot \left( \int_a^x p(\omega) (|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p) d\omega \right)^{\left(\frac{\lambda_\alpha + \lambda_\beta + \lambda_\nu}{p}\right)}, \end{aligned} \quad (48)$$

all  $a \leq x \leq b$ .

**Proof.** As Theorem 4 of [3], and Theorem 6 of [9].  $\square$

We continue with the special important case

**Theorem 33.** *Let  $\nu \geq \gamma_1 + 2$ ,  $\gamma_1 \geq 0$ ,  $n := \lceil \nu \rceil$  and  $f_1, f_2 \in AC^n([a, b])$  such that  $f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ . Consider also  $p(t) > 0$  and  $q(t) \geq 0$ , with all  $p(t)$ ,  $\frac{1}{p(t)}$ ,  $q(t) \in L_\infty(a, b)$ . Furthermore assume  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ .*

*Let  $\lambda_\alpha \geq 0$ ,  $0 < \lambda_{\alpha+1} < 1$ , and  $p > 1$ . Denote*

$$\theta_3 := \begin{cases} 2^{\lambda_\alpha/\lambda_{\alpha+1}} - 1, & \text{if } \lambda_\alpha \geq \lambda_{\alpha+1}, \\ 1, & \text{if } \lambda_\alpha \leq \lambda_{\alpha+1}, \end{cases} \quad (49)$$

$$L(x) := \left( 2 \int_a^x (q(\omega))^{\left(\frac{1}{1-\lambda_{\alpha+1}}\right)} d\omega \right)^{(1-\lambda_{\alpha+1})} \left( \frac{\theta_3 \lambda_{\alpha+1}}{\lambda_\alpha + \lambda_{\alpha+1}} \right)^{\lambda_{\alpha+1}}, \quad (50)$$

and

$$P_1(x) := \int_a^x (x-t)^{\frac{(\nu-\gamma_1-1)p}{p-1}} (p(t))^{-1/(p-1)} dt, \quad (51)$$

$$T(x) := L(x) \left( \frac{P_1(x)^{\left(\frac{p-1}{p}\right)}}{\Gamma(\nu - \gamma_1)} \right)^{(\lambda_\alpha + \lambda_{\alpha+1})}, \quad (52)$$

$$\omega_1 := 2^{\left(\frac{p-1}{p}\right)(\lambda_\alpha + \lambda_{\alpha+1})}, \quad (53)$$

and

$$\Phi(x) := T(x) \omega_1. \quad (54)$$

Then

$$\begin{aligned} & \int_a^x q(\omega) \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_2(\omega)|^{\lambda_{\alpha+1}} + \right. \\ & \quad \left. |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_1(\omega)|^{\lambda_{\alpha+1}} \right] d\omega \leq \\ & \quad \Phi(x) \left[ \int_a^x p(\omega) (|D_{*a}^\nu f_1(\omega)|^p + \right. \\ & \quad \left. |D_{*a}^\nu f_2(\omega)|^p) d\omega \right]^{\left( \frac{\lambda_\alpha + \lambda_{\alpha+1}}{p} \right)}, \end{aligned} \quad (55)$$

all  $a \leq x \leq b$ .

**Proof.** As in Theorem 5 of [3], and Theorem 8 of [9].  $\square$

We give

**Corollary 34.** All here as in Theorem 30, with  $\lambda_\beta = 0$ ,  $p(t) = q(t) = 1$ .

Then

$$\begin{aligned} & \int_a^x \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \quad \left. |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \leq \\ & \quad C_1(x) \left( \int_a^x (|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p) d\omega \right)^{\left( \frac{\lambda_\alpha + \lambda_\nu}{p} \right)}, \end{aligned} \quad (56)$$

all  $a \leq x \leq b$ , where

$$C_1(x) := (A_0(x)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{\lambda_\alpha + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_1, \quad (57)$$

$$\delta_1 := \begin{cases} 2^{1 - \left( \frac{\lambda_\alpha + \lambda_\nu}{p} \right)}, & \text{if } \lambda_\alpha + \lambda_\nu \leq p, \\ 1, & \text{if } \lambda_\alpha + \lambda_\nu \geq p. \end{cases} \quad (58)$$

We find that

$$\begin{aligned} (A_0(x)|_{\lambda_\beta=0}) &= \left\{ \left( \frac{(p-1)^{\left( \frac{\lambda_\alpha p - \lambda_\alpha}{p} \right)}}{(\Gamma(\nu - \gamma_1))^{\lambda_\alpha} (\nu p - \gamma_1 p - 1)^{\left( \frac{\lambda_\alpha p - \lambda_\alpha}{p} \right)}} \right) \cdot \right. \\ & \quad \left. \left( \frac{(p - \lambda_\nu)^{\left( \frac{p - \lambda_\nu}{p} \right)}}{(\lambda_\alpha \nu p - \lambda_\alpha \gamma_1 p - \lambda_\alpha + p - \lambda_\nu)^{\left( \frac{p - \lambda_\nu}{p} \right)}} \right) \right\} \cdot \\ & \quad (x - a)^{\left( \frac{\lambda_\alpha \nu p - \lambda_\alpha \gamma_1 p - \lambda_\alpha + p - \lambda_\nu}{p} \right)}. \end{aligned} \quad (59)$$

**Proof.** As Corollary 1 of [3], and Corollary 10 of [9].  $\square$

**Corollary 35.** (All as in Theorem 30,  $\lambda_\beta = 0$ ,  $p(t) = q(t) = 1$ ,  $\lambda_\alpha = \lambda_\nu = 1$ ,  $p = 2$ .)

*In detail:* Let  $\nu \geq \gamma_1 + 1$ ,  $\gamma_1 \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f_1, f_2 \in AC^n([a, b]) : f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ ,  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ . Then

$$\int_a^x [| (D_{*a}^{\gamma_1} f_1)(\omega) | | (D_{*a}^\nu f_1)(\omega) | + | (D_{*a}^{\gamma_1} f_2)(\omega) | | (D_{*a}^\nu f_2)(\omega) |] d\omega \leq \quad (60)$$

$$\left( \frac{(x-a)^{(\nu-\gamma_1)}}{2\Gamma(\nu-\gamma_1)\sqrt{\nu-\gamma_1}\sqrt{2\nu-2\gamma_1-1}} \right) \left( \int_a^x [((D_{*a}^\nu f_1)(\omega))^2 + ((D_{*a}^\nu f_2)(\omega))^2] d\omega \right),$$

all  $a \leq x \leq b$ .

**Corollary 36.** (to Theorem 31;  $\lambda_\alpha = 0$ ,  $p(t) = q(t) = 1$ .)

*It holds*

$$\int_a^x [|D_{*a}^{\gamma_2} f_2(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + |D_{*a}^{\gamma_2} f_1(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu}] d\omega \leq \quad (61)$$

$$C_2(x) \left( \int_a^x [|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p] d\omega \right)^{\left(\frac{\lambda_\nu + \lambda_\beta}{p}\right)},$$

all  $a \leq x \leq b$ , where

$$C_2(x) := (A_0(x)|_{\lambda_\alpha=0}) 2^{\frac{p-\lambda_\nu}{p}} \left( \frac{\lambda_\nu}{\lambda_\beta + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_3^{\lambda_\nu/p}, \quad (62)$$

$$\delta_3 := \begin{cases} 2^{\lambda_\beta/\lambda_\nu} - 1, & \text{if } \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\beta \leq \lambda_\nu \end{cases}. \quad (63)$$

We find that

$$(A_0(x)|_{\lambda_\alpha=0}) = \left\{ \left( \frac{(p-1)^{\left(\frac{\lambda_\beta p - \lambda_\beta}{p}\right)}}{(\Gamma(\nu-\gamma_2))^{\lambda_\beta} (\nu p - \gamma_2 p - 1)^{\left(\frac{\lambda_\beta p - \lambda_\beta}{p}\right)}} \right) \cdot \left( \frac{(p-\lambda_\nu)^{\left(\frac{p-\lambda_\nu}{p}\right)}}{(\lambda_\beta \nu p - \lambda_\beta \gamma_2 p - \lambda_\beta + p - \lambda_\nu)^{\left(\frac{p-\lambda_\nu}{p}\right)}} \right) \right\} \cdot (x-a)^{\left(\frac{\lambda_\beta \nu p - \lambda_\beta \gamma_2 p - \lambda_\beta + p - \lambda_\nu}{p}\right)}. \quad (64)$$

**Corollary 37.** (to Theorem 31;  $\lambda_\alpha = 0$ ,  $p(t) = q(t) = 1$ ,  $\lambda_\beta = \lambda_\nu = 1$ ,  $p = 2$ .)  
In detail:

Let  $\nu \geq \gamma_2 + 1$ ,  $\gamma_2 \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f_1, f_2 \in AC^n([a, b]) : f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ ,  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ .

Then

$$\begin{aligned} & \int_a^x [|D_{*a}^{\gamma_2} f_2(\omega)| |D_{*a}^\nu f_1(\omega)| + \\ & |D_{*a}^{\gamma_2} f_1(\omega)| |D_{*a}^\nu f_2(\omega)|] d\omega \leq \\ & \left( \frac{(x-a)^{(\nu-\gamma_2)}}{\sqrt{2}\Gamma(\nu-\gamma_2)\sqrt{\nu-\gamma_2}\sqrt{2\nu-2\gamma_2-1}} \right) \cdot \\ & \left( \int_a^x \left( (D_{*a}^\nu f_1(\omega))^2 + (D_{*a}^\nu f_2(\omega))^2 \right) d\omega \right), \end{aligned} \quad (65)$$

all  $a \leq x \leq b$ .

We continue with related results regarding  $\|\cdot\|_\infty$ .

**Theorem 38.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, 2$ ,  $n := \lceil \nu \rceil$  and  $f_1, f_2 \in AC^n([a, b])$  such that  $f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ;  $a \leq x \leq b$ . Consider  $p(x) \geq 0$  and  $p(x) \in L_\infty(a, b)$ , and assume  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ . Let  $\lambda_\alpha, \lambda_\beta, \lambda_\nu \geq 0$ . Set

$$\begin{aligned} T(x) &:= \frac{(x-a)^{(\nu\lambda_\alpha - \gamma_1\lambda_\alpha + \nu\lambda_\beta - \gamma_2\lambda_\beta + 1)}}{(\nu\lambda_\alpha - \gamma_1\lambda_\alpha + \nu\lambda_\beta - \gamma_2\lambda_\beta + 1)} \\ &\cdot \frac{\|p(s)\|_{\infty, (a, x)}}{(\Gamma(\nu - \gamma_1 + 1))^{\lambda_\alpha} (\Gamma(\nu - \gamma_2 + 1))^{\lambda_\beta}}. \end{aligned} \quad (66)$$

Then

$$\begin{aligned} & \int_a^x p(\omega) \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_2} f_2(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \left. |D_{*a}^{\gamma_2} f_1(\omega)|^{\lambda_\beta} |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \leq \\ & \frac{T(x)}{2} \left[ \|D_{*a}^\nu f_1\|_{\infty, (a, x)}^{2(\lambda_\alpha + \lambda_\nu)} + \|D_{*a}^\nu f_1\|_{\infty, (a, x)}^{2\lambda_\beta} + \right. \\ & \left. \|D_{*a}^\nu f_2\|_{\infty, (a, x)}^{2\lambda_\beta} + \|D_{*a}^\nu f_2\|_{\infty, (a, x)}^{2(\lambda_\alpha + \lambda_\nu)} \right], \end{aligned} \quad (67)$$

all  $a \leq x \leq b$ .

**Proof.** Similar to Theorem 7 of [3], and Theorem 18 of [9].  $\square$

We give

**Corollary 39.** (to Theorem 38) Let  $\nu \geq \gamma_1 + 2$ ,  $\gamma_1 \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f_1, f_2 \in AC^n([a, b])$  such that  $f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ;  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ . Then

$$\begin{aligned}
& \int_a^x [|D_{*a}^{\gamma_1} f_1(\omega)| |D_{*a}^{\gamma_1+1} f_2(\omega)| + \\
& |D_{*a}^{\gamma_1+1} f_1(\omega)| |D_{*a}^{\gamma_1} f_2(\omega)|] d\omega \\
& \leq \frac{(x-a)^{2(\nu-\gamma_1)}}{2(\Gamma(\nu-\gamma_1+1))^2} \\
& \left[ \|D_{*a}^\nu f_1\|_{\infty, (a,x)}^2 + \|D_{*a}^\nu f_2\|_{\infty, (a,x)}^2 \right], \tag{68}
\end{aligned}$$

all  $a \leq x \leq b$ .

Next we give converse results involving two functions.

**Theorem 40.** Let  $\gamma_j \geq 0, 1 \leq \nu - \gamma_j < \frac{1}{p}, 0 < p < 1, j = 1, 2; n := \lceil \nu \rceil$ , and  $f_1, f_2 \in AC^n([a, b])$  such that  $f_1^{(l)}(a) = f_2^{(l)}(a) = 0, l = 0, 1, \dots, n-1, a \leq x \leq b$ . Consider also  $p(t) > 0$  and  $q(t) \geq 0$ , with all  $p(t), \frac{1}{p(t)}, q(t), \frac{1}{q(t)} \in L_\infty(a, b)$ . Further assume  $D_{*a}^\nu f_i \in L_\infty(a, b), i = 1, 2$ ; each of which is of fixed sign a.e. on  $[a, b]$ . Let  $\lambda_\nu > 0$  and  $\lambda_\alpha, \lambda_\beta \geq 0$  such that  $\lambda_\nu > p$ .

Here  $P_k(\omega), A(\omega), A_0(x)$  are as in (39), (40), (41), respectively.

Set

$$\delta_1 := 2^{1 - (\frac{\lambda_\alpha + \lambda_\nu}{p})}. \tag{69}$$

If  $\lambda_\beta = 0$ , then

$$\begin{aligned}
& \int_a^x q(\omega) \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\
& \left. |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \geq \\
& (A_0(x)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{\lambda_\alpha + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_1 \\
& \left[ \int_a^x p(\omega) [|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p] d\omega \right]^{\left( \frac{\lambda_\alpha + \lambda_\nu}{p} \right)}. \tag{70}
\end{aligned}$$

**Proof.** Similar to Theorem 5 of [8] and Theorem 4 of [7].  $\square$

We continue with

**Theorem 41.** All here as in Theorem 40. Further assume  $\lambda_\beta \geq \lambda_\nu$ . Denote

$$\delta_2 := 2^{1 - (\lambda_\beta/\lambda_\nu)},$$

$$\delta_3 := (\delta_2 - 1) 2^{-(\lambda_\beta/\lambda_\nu)}. \tag{71}$$

If  $\lambda_\alpha = 0$ , then

$$\int_a^x q(\omega) \left[ |D_{*a}^{\gamma_2} f_2(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right.$$

$$\begin{aligned}
& |D_{*a}^{\gamma_2} f_1(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} d\omega \geq \\
& (A_0(x)|_{\lambda_\alpha=0}) 2^{\frac{p-\lambda_\nu}{p}} \left( \frac{\lambda_\nu}{\lambda_\beta + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_3^{\lambda_\nu/p} \\
& \cdot \left( \int_a^x p(\omega) [|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p] d\omega \right)^{\left( \frac{\lambda_\nu + \lambda_\beta}{p} \right)}, \quad (72)
\end{aligned}$$

all  $a \leq x \leq b$ .

**Proof.** Similar to Theorem 6 of [8], and Theorem 5 of [7].  $\square$

We give

**Theorem 42.** Let  $\nu \geq 2$  and  $\gamma_1 \geq 0$  such that  $2 \leq \nu - \gamma_1 < \frac{1}{p}$ ,  $0 < p < 1$ ,  $n := \lceil \nu \rceil$ . Consider  $f_1, f_2 \in AC^n([a, b])$  such that  $f_1^{(l)}(a) = f_2^{(l)}(a) = 0$ ,  $l = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ . Assume that  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ ; each of which is of fixed sign a.e. on  $[a, b]$ . Consider also  $p(t) > 0$  and  $q(t) \geq 0$ , with all  $p(t)$ ,  $\frac{1}{p(t)}$ ,  $q(t)$ ,  $\frac{1}{q(t)} \in L_\infty(a, b)$ . Let  $\lambda_\alpha \geq \lambda_{\alpha+1} > 1$ . Denote

$$\theta_3 := \left( 2^{1-(\lambda_\alpha/\lambda_{\alpha+1})} - 1 \right) 2^{-\lambda_\alpha/\lambda_{\alpha+1}}, \quad (73)$$

$L(x)$  as in (50),  $P_1(x)$  as in (51),  $T(x)$  as in (52),  $\omega_1$  as in (53), and  $\Phi$  as in (54). Then

$$\begin{aligned}
& \int_a^x q(\omega) \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_2(\omega)|^{\lambda_{\alpha+1}} + \right. \\
& \left. |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_1(\omega)|^{\lambda_{\alpha+1}} \right] d\omega \geq \\
& \Phi(x) \left[ \int_a^x p(\omega) (|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p) d\omega \right]^{\left( \frac{\lambda_\alpha + \lambda_{\alpha+1}}{p} \right)}, \quad (74)
\end{aligned}$$

all  $a \leq x \leq b$ .

**Proof.** Similar to Theorem 7 of [8], and Theorem 7 of [7].  $\square$

We have

**Corollary 43.** (to Theorem 40;  $\lambda_\beta = 0$ ,  $p(t) = q(t) = 1$ ). Then

$$\begin{aligned}
& \int_a^x \left[ |D_{*a}^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\
& \left. |D_{*a}^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \geq \\
& C_1(x) \left( \int_a^x (|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p) d\omega \right)^{\left( \frac{\lambda_\alpha + \lambda_\nu}{p} \right)}, \quad (75)
\end{aligned}$$

all  $a \leq x \leq b$ , where

$$C_1(x) := (A_0(x)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{\lambda_\alpha + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_1, \quad (76)$$

$$\delta_1 := 2^{1-(\frac{\lambda_\alpha+\lambda_\nu}{p})}. \quad (77)$$

Here  $(A_0(x)|_{\lambda_\beta=0})$  is given by (59).

**Corollary 44.** (to Theorem 41;  $\lambda_\alpha = 0$ ,  $p(t) = q(t) = 1$ ,  $\lambda_\beta \geq \lambda_\nu$ ). Then

$$\begin{aligned} & \int_a^x \left[ |D_{*a}^{\gamma_2} f_2(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \left. |D_{*a}^{\gamma_2} f_1(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \geq \\ & C_2(x) \left( \int_a^x [|D_{*a}^\nu f_1(\omega)|^p + |D_{*a}^\nu f_2(\omega)|^p] d\omega \right)^{\left( \frac{\lambda_\nu + \lambda_\beta}{p} \right)}, \end{aligned} \quad (78)$$

all  $a \leq x \leq b$ , where

$$C_2(x) := (A_0(x)|_{\lambda_\alpha=0}) 2^{\frac{p-\lambda_\nu}{p}} \left( \frac{\lambda_\nu}{\lambda_\beta + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_3^{\lambda_\nu/p}.$$

Here  $(A_0(x)|_{\lambda_\alpha=0})$  is given by (64).

### 3.3 Results involving several functions

We present

**Theorem 45.** Here all notations, terms and assumptions are as in Theorem 30, but for  $f_j \in AC^n([a, b])$ , with  $j = 1, \dots, M \in \mathbb{N}$ . Instead of  $\delta_1$  there, we define here

$$\delta_1^* := \begin{cases} M^{1-(\frac{\lambda_\alpha+\lambda_\nu}{p})}, & \text{if } \lambda_\alpha + \lambda_\nu \leq p, \\ 2^{(\frac{\lambda_\alpha+\lambda_\nu}{p})-1}, & \text{if } \lambda_\alpha + \lambda_\nu \geq p. \end{cases} \quad (79)$$

Call

$$\varphi_1(x) := (A_0(x)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{\lambda_\alpha + \lambda_\nu} \right)^{\lambda_\nu/p}. \quad (80)$$

If  $\lambda_\beta = 0$ , then

$$\begin{aligned} & \int_a^x q(\omega) \left( \sum_{j=1}^M |D_{*a}^{\gamma_1} f_j(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_j(\omega)|^{\lambda_\nu} \right) d\omega \\ & \leq \delta_1^* \varphi_1(x) \left[ \int_a^x p(\omega) \left( \sum_{j=1}^M |D_{*a}^\nu f_j(\omega)|^p \right) d\omega \right]^{\left( \frac{\lambda_\alpha + \lambda_\nu}{p} \right)}, \end{aligned} \quad (81)$$

all  $a \leq x \leq b$ .

**Proof.** As in Theorem 2 of [4], and Theorem 4 of [10].  $\square$

We continue with

**Theorem 46.** All here as in Theorem 45.

Denote

$$\delta_3 := \begin{cases} 2^{\lambda_\beta/\lambda_\nu} - 1, & \text{if } \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\beta \leq \lambda_\nu, \end{cases} \quad (82)$$

$$\varepsilon_2 := \begin{cases} 1, & \text{if } \lambda_\nu + \lambda_\beta \geq p, \\ M^{1 - \left(\frac{\lambda_\nu + \lambda_\beta}{p}\right)}, & \text{if } \lambda_\nu + \lambda_\beta \leq p, \end{cases} \quad (83)$$

and

$$\varphi_2(x) := (A_0(x)|_{\lambda_\alpha=0}) 2^{\left(\frac{p-\lambda_\nu}{p}\right)} \left(\frac{\lambda_\nu}{\lambda_\beta + \lambda_\nu}\right)^{\lambda_\nu/p} \delta_3^{\lambda_\nu/p}. \quad (84)$$

If  $\lambda_\alpha = 0$ , then

$$\begin{aligned} \int_a^x q(\omega) & \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |D_{*a}^{\gamma_2} f_{j+1}(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_j(\omega)|^{\lambda_\nu} + \right. \right. \right. \\ & \left. \left. |D_{*a}^{\gamma_2} f_j(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_{j+1}(\omega)|^{\lambda_\nu} \right] \right\} + \\ & \left[ |D_{*a}^{\gamma_2} f_M(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \left. |D_{*a}^{\gamma_2} f_1(\omega)|^{\lambda_\beta} |D_{*a}^\nu f_M(\omega)|^{\lambda_\nu} \right] \Big\} d\omega \\ & \leq 2^{\left(\frac{\lambda_\nu + \lambda_\beta}{p}\right)} \varepsilon_2 \varphi_2(x) \cdot \left\{ \int_a^x p(\omega) \right. \\ & \left. \cdot \left[ \sum_{j=1}^M |D_{*a}^\nu f_j(\omega)|^p \right] d\omega \right\}^{\left(\frac{\lambda_\nu + \lambda_\beta}{p}\right)}, \end{aligned} \quad (85)$$

all  $a \leq x \leq b$ .

**Proof.** As Theorem 3 of [4], and Theorem 5 of [10].  $\square$

We give the general case

**Theorem 47.** All as in Theorem 45.

Denote

$$\tilde{\gamma}_1 := \begin{cases} 2^{\left(\frac{\lambda_\alpha + \lambda_\beta}{\lambda_\nu}\right)} - 1, & \text{if } \lambda_\alpha + \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\alpha + \lambda_\beta \leq \lambda_\nu, \end{cases} \quad (86)$$

and

$$\tilde{\gamma}_2 := \begin{cases} 1, & \text{if } \lambda_\alpha + \lambda_\beta + \lambda_\nu \geq p, \\ 2^{1 - \left(\frac{\lambda_\alpha + \lambda_\beta + \lambda_\nu}{p}\right)}, & \text{if } \lambda_\alpha + \lambda_\beta + \lambda_\nu \leq p. \end{cases} \quad (87)$$



Set

$$\begin{aligned} \varphi_3(x) := A_0(x) \cdot \left( \frac{\lambda_\nu}{(\lambda_\alpha + \lambda_\beta)(\lambda_\alpha + \lambda_\beta + \lambda_\nu)} \right)^{\frac{\lambda_\nu}{p}} \\ \cdot \left[ \lambda_\alpha^{\frac{\lambda_\nu}{p}} \tilde{\gamma}_2 + 2^{\left(\frac{p-\lambda_\nu}{p}\right)} (\tilde{\gamma}_1 \lambda_\beta)^{\frac{\lambda_\nu}{p}} \right], \end{aligned} \quad (88)$$

and

$$\varepsilon_3 := \begin{cases} 1, & \text{if } \lambda_\alpha + \lambda_\beta + \lambda_\nu \geq p, \\ M^{1-\left(\frac{\lambda_\alpha + \lambda_\beta + \lambda_\nu}{p}\right)} & \text{if } \lambda_\alpha + \lambda_\beta + \lambda_\nu \leq p. \end{cases} \quad (89)$$

Then

$$\begin{aligned} \int_a^x q(\omega) \left[ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_1} f_j)(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_2} f_{j+1})(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_j)(\omega)|^{\lambda_\nu} \right. \right. \\ \left. \left. + |(D_{*a}^{\gamma_2} f_j)(\omega)|^{\lambda_\beta} |(D_{*a}^{\gamma_1} f_{j+1})(\omega)|^{\lambda_\alpha} |(D_{*a}^\nu f_{j+1})(\omega)|^{\lambda_\nu} \right] \right. \\ \left. + \left[ |(D_{*a}^{\gamma_1} f_1)(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_2} f_M)(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_1)(\omega)|^{\lambda_\nu} \right. \right. \\ \left. \left. + |(D_{*a}^{\gamma_2} f_1)(\omega)|^{\lambda_\beta} |(D_{*a}^{\gamma_1} f_M)(\omega)|^{\lambda_\alpha} |(D_{*a}^\nu f_M)(\omega)|^{\lambda_\nu} \right] \right] d\omega \\ \leq 2^{\left(\frac{\lambda_\alpha + \lambda_\beta + \lambda_\nu}{p}\right)} \varepsilon_3 \varphi_3(x) \cdot \left\{ \int_a^x p(\omega) \right. \\ \left. \cdot \left[ \sum_{j=1}^M |(D_{*a}^\nu f_j)(\omega)|^p \right] d\omega \right\}^{\left(\frac{\lambda_\alpha + \lambda_\beta + \lambda_\nu}{p}\right)}, \end{aligned} \quad (90)$$

all  $a \leq x \leq b$ .

**Proof.** As Theorem 4 of [4], and Theorem 6 of [10].  $\square$

We give

**Theorem 48.** All here as in Theorem 33, but for  $f_j \in AC^n([a, b])$ ,  $j = 1, \dots, M \in \mathbb{N}$ .

Also put

$$\varepsilon_4 := \begin{cases} 1, & \text{if } \lambda_\alpha + \lambda_{\alpha+1} \geq p, \\ M^{1-\left(\frac{\lambda_\alpha + \lambda_{\alpha+1}}{p}\right)} & \text{if } \lambda_\alpha + \lambda_{\alpha+1} \leq p. \end{cases} \quad (91)$$

Then

$$\int_a^x q(\omega) \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_1} f_j)(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_1+1} f_{j+1})(\omega)|^{\lambda_{\alpha+1}} \right. \right. \right.$$

$$\begin{aligned}
& + |(D_{*a}^{\gamma_1} f_{j+1})(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_1+1} f_j)(\omega)|^{\lambda_{\alpha+1}} \Big] \Big\} \\
& + \Big[ |(D_{*a}^{\gamma_1} f_1)(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_1+1} f_M)(\omega)|^{\lambda_{\alpha+1}} \\
& + |(D_{*a}^{\gamma_1} f_M)(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_1+1} f_1)(\omega)|^{\lambda_{\alpha+1}} \Big] d\omega \\
& \leq 2^{\left(\frac{\lambda_\alpha + \lambda_{\alpha+1}}{p}\right)} \varepsilon_4 \Phi(x) \cdot \left[ \int_a^x p(\omega) \right. \\
& \quad \cdot \left. \left( \sum_{j=1}^M |(D_{*a}^\nu f_j)(\omega)|^p \right) d\omega \right]^{\left(\frac{\lambda_\alpha + \lambda_{\alpha+1}}{p}\right)}, \tag{92}
\end{aligned}$$

all  $a \leq x \leq b$ .

**Proof.** As Theorem 5 of [4], and Theorem 7 of [10].  $\square$

We continue with

**Corollary 49.** (to Theorem 45,  $\lambda_\beta = 0$ ,  $p(t) = q(t) = 1$ ,  $\lambda_\alpha = \lambda_\nu = 1$ ,  $p = 2$ ). *In detail:*

Let  $\nu \geq \gamma_1 + 1$ ,  $\gamma_1 \geq 0$ ,  $n := [\nu]$ ,  $f_j \in AC^n([a, b])$ ,  $j = 1, \dots, M \in \mathbb{N}$ ;  $a \leq x \leq b$ , and  $D_{*a}^\nu f_j \in L_\infty(a, b)$ ,  $j = 1, \dots, M$ .

Here  $f_j^{(l)}(a) = 0$ ,  $l = 0, 1, \dots, n-1$ ;  $j = 1, \dots, M$ .

Then

$$\begin{aligned}
& \int_a^x \left( \sum_{j=1}^M |D_{*a}^{\gamma_1} f_j(\omega)| |D_{*a}^\nu f_j(\omega)| \right) d\omega \leq \\
& \left( \frac{(x-a)^{\nu-\gamma_1}}{2\Gamma(\nu-\gamma_1) \sqrt{\nu-\gamma_1} \sqrt{2\nu-2\gamma_1-1}} \right) \cdot \\
& \left\{ \int_a^x \left[ \sum_{j=1}^M (D_{*a}^\nu f_j(\omega))^2 \right] d\omega \right\}, \tag{93}
\end{aligned}$$

all  $a \leq x \leq b$ .

**Corollary 50.** (to Theorem 46,  $\lambda_\alpha = 0$ ,  $p(t) = q(t) = 1$ ,  $\lambda_\beta = \lambda_\nu = 1$ ,  $p = 2$ ). *In detail:*

Let  $\nu \geq \gamma_2 + 1$ ,  $\gamma_2 \geq 0$ ,  $n := [\nu]$ ,  $f_j \in AC^n([a, b])$ ,  $D_{*a}^\nu f_j \in L_\infty(a, b)$ ,  $j = 1, \dots, M \in \mathbb{N}$ ;  $a \leq x \leq b$ .

Here  $f_j^{(l)}(a) = 0$ ,  $l = 0, 1, \dots, n-1$ ;  $j = 1, \dots, M$ .

Then

$$\int_a^x \left\{ \left\{ \sum_{j=1}^{M-1} [| (D_{*a}^{\gamma_2} f_{j+1})(\omega) | | (D_{*a}^\nu f_j)(\omega) | \right. \right.$$

$$\begin{aligned}
& + |(D_{*a}^{\gamma_2} f_j)(\omega)| |(D_{*a}^{\nu} f_{j+1})(\omega)| \} \\
& + [| (D_{*a}^{\gamma_2} f_M)(\omega)| |(D_{*a}^{\nu} f_1)(\omega)| \\
& + |(D_{*a}^{\gamma_2} f_1)(\omega)| |(D_{*a}^{\nu} f_M)(\omega)| \} d\omega \\
& \leq \left( \frac{\sqrt{2}(x-a)^{(\nu-\gamma_2)}}{\Gamma(\nu-\gamma_2)\sqrt{\nu-\gamma_2}\sqrt{2\nu-2\gamma_2-1}} \right) \\
& \quad \left\{ \int_a^x \left[ \sum_{j=1}^M ((D_{*a}^{\nu} f_j)(\omega))^2 \right] d\omega \right\}, \tag{94}
\end{aligned}$$

all  $a \leq x \leq b$ .

**Corollary 51.** (to Theorem 48,  $\lambda_\alpha = 1$ ,  $\lambda_{\alpha+1} = 1/2$ ,  $p = 3/2$ ,  $p(t) = q(t) = 1$ ). In detail:

Let  $\nu \geq \gamma_1 + 2$ ,  $\gamma_1 \geq 0$ ,  $n := \lceil \nu \rceil$ , and  $f_j \in AC^n([a, b])$ ,  $j = 1, \dots, M \in \mathbb{N}$ , such that  $f_j^{(l)}(a) = 0$ ,  $l = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ . Assume also  $D_{*a}^{\nu} f_j \in L_\infty(a, b)$ ,  $j = 1, \dots, M$ .

Set

$$\Phi^*(x) := \left( \frac{2}{\sqrt{3\nu-3\gamma_1-2}} \right) \frac{(x-a)^{\left(\frac{3\nu-3\gamma_1-1}{2}\right)}}{(\Gamma(\nu-\gamma_1))^{3/2}}, \tag{95}$$

all  $a \leq x \leq b$ . Then

$$\begin{aligned}
& \int_a^x \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_1} f_j)(\omega)| \sqrt{|(D_{*a}^{\gamma_1+1} f_{j+1})(\omega)|} \right. \right. \right. \\
& \quad \left. \left. \left. + |(D_{*a}^{\gamma_1} f_{j+1})(\omega)| \sqrt{|(D_{*a}^{\gamma_1+1} f_j)(\omega)|} \right] \right\} \right. \\
& \quad \left. + \left[ |(D_{*a}^{\gamma_1} f_1)(\omega)| \sqrt{|(D_{*a}^{\gamma_1+1} f_M)(\omega)|} \right. \right. \\
& \quad \left. \left. + |(D_{*a}^{\gamma_1} f_M)(\omega)| \sqrt{|(D_{*a}^{\gamma_1+1} f_1)(\omega)|} \right] \right\} d\omega \\
& \leq 2\Phi^*(x) \cdot \left[ \int_a^x \left( \sum_{j=1}^M |(D_{*a}^{\nu} f_j)(\omega)|^{3/2} \right) d\omega \right], \tag{96}
\end{aligned}$$

all  $a \leq x \leq b$ .

We continue with results regarding  $\|\cdot\|_\infty$ .

**Theorem 52.** All as in Theorem 38 but for  $f_j \in AC^n([a, b])$ ,  $j = 1, \dots, M \in \mathbb{N}$ . Then

$$\begin{aligned}
& \int_a^x p(\omega) \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_1} f_j)(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_2} f_{j+1})(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_j)(\omega)|^{\lambda_\nu} \right. \right. \right. \\
& \quad \left. \left. \left. + |(D_{*a}^{\gamma_2} f_j)(\omega)|^{\lambda_\beta} |(D_{*a}^{\gamma_1} f_{j+1})(\omega)|^{\lambda_\alpha} |(D_{*a}^\nu f_{j+1})(\omega)|^{\lambda_\nu} \right] \right\} \right. \\
& \quad \left. + \left[ |(D_{*a}^{\gamma_1} f_1)(\omega)|^{\lambda_\alpha} |(D_{*a}^{\gamma_2} f_M)(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_1)(\omega)|^{\lambda_\nu} \right. \right. \\
& \quad \left. \left. + |(D_{*a}^{\gamma_2} f_1)(\omega)|^{\lambda_\beta} |(D_{*a}^{\gamma_1} f_M)(\omega)|^{\lambda_\alpha} |(D_{*a}^\nu f_M)(\omega)|^{\lambda_\nu} \right] \right\} d\omega \\
& \leq T(x) \left\{ \sum_{j=1}^M \left\{ \|D_{*a}^\nu f_j\|_{\infty, (a, x)}^{2(\lambda_\alpha + \lambda_\nu)} + \|D_{*a}^\nu f_j\|_{\infty, (a, x)}^{2\lambda_\beta} \right\} \right\}, \quad (97)
\end{aligned}$$

all  $a \leq x \leq b$ .

**Proof.** Based on Theorem 38.  $\square$

We give

**Corollary 53.** (to Theorem 52) *In detail:*

Let  $\nu \geq \gamma_1 + 2$ ,  $\gamma_1 \geq 0$ ,  $n := \lceil \nu \rceil$  and  $f_j \in AC^n([a, b])$ ,  $j = 1, \dots, M \in \mathbb{N}$ , such that  $f_j^{(l)}(a) = 0$ ,  $l = 0, 1, \dots, n-1$ ;  $j = 1, \dots, M$ ;  $a \leq x \leq b$ . Further suppose that  $D_{*a}^\nu f_j \in L_\infty(a, b)$ ,  $j = 1, \dots, M$ . Then

$$\begin{aligned}
& \int_a^x \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_1} f_j)(\omega)| |(D_{*a}^{\gamma_1+1} f_{j+1})(\omega)| \right. \right. \right. \\
& \quad \left. \left. \left. + |(D_{*a}^{\gamma_1+1} f_j)(\omega)| |(D_{*a}^{\gamma_1} f_{j+1})(\omega)| \right] \right\} \right. \\
& \quad \left. + \left[ |(D_{*a}^{\gamma_1} f_1)(\omega)| |(D_{*a}^{\gamma_1+1} f_M)(\omega)| \right. \right. \\
& \quad \left. \left. + |(D_{*a}^{\gamma_1+1} f_1)(\omega)| |(D_{*a}^{\gamma_1} f_M)(\omega)| \right] \right\} d\omega \\
& \leq \left( \frac{(x-a)^{2(\nu-\gamma_1)}}{(\Gamma(\nu-\gamma_1+1))^2} \right) \left( \sum_{j=1}^M \|D_{*a}^\nu f_j\|_\infty^2 \right), \quad (98)
\end{aligned}$$

all  $a \leq x \leq b$ .

We continue with converse results.

**Theorem 54.** Let  $\gamma_j \geq 0$ ,  $1 \leq \nu - \gamma_j < \frac{1}{p}$ ,  $0 < p < 1$ ,  $j = 1, 2$ ;  $n := \lceil \nu \rceil$ , and  $f_i \in AC^n([a, b])$ ,  $i = 1, \dots, M \in \mathbb{N}$ , such that  $f_i^{(l)}(a) = 0$ ,  $l = 0, 1, \dots, n-1$ ;  $i = 1, \dots, M$ ;  $a \leq x \leq b$ . Consider also  $p(t) > 0$  and  $q(t) \geq 0$ , with all  $p(t)$ ,  $\frac{1}{p(t)}$ ,  $q(t)$ ,  $\frac{1}{q(t)} \in L_\infty(a, b)$ . Further assume  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, \dots, M$ ; each of which is of fixed sign a.e. on  $[a, b]$ . Let  $\lambda_\nu > 0$  and  $\lambda_\alpha, \lambda_\beta \geq 0$  such that  $\lambda_\nu > p$ .

Here  $P_k(\omega)$ ,  $k = 1, 2$ ,  $A(\omega)$ ,  $A_0(x)$  are as in (39), (40), (41), respectively. Call

$$\varphi_1(x) := (A_0(x)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{\lambda_\alpha + \lambda_\nu} \right)^{\lambda_\nu/p}, \quad (99)$$

$$\delta_1^* := M^{1-(\frac{\lambda_\alpha+\lambda_\nu}{p})}. \quad (100)$$

If  $\lambda_\beta = 0$ , then

$$\begin{aligned} & \int_a^x q(\omega) \left( \sum_{j=1}^M |D_{*a}^{\gamma_1} f_j(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_j(\omega)|^{\lambda_\nu} \right) d\omega \\ & \geq \delta_1^* \varphi_1(x) \left[ \int_a^x p(\omega) \left( \sum_{j=1}^M |D_{*a}^\nu f_j(\omega)|^p \right) d\omega \right]^{\left(\frac{\lambda_\alpha+\lambda_\nu}{p}\right)}, \end{aligned} \quad (101)$$

all  $a \leq x \leq b$ .

**Proof.** As Theorem 11 of [7], and Theorem 11 of [8].  $\square$

We continue with

**Theorem 55.** All as in Theorem 54. Assume  $\lambda_\beta \geq \lambda_\nu$ . Denote

$$\varphi_2(x) := (A_0(x)|_{\lambda_\alpha=0}) 2^{\left(\frac{p-\lambda_\nu}{p}\right)} \left( \frac{\lambda_\nu}{\lambda_\beta + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_3^{\lambda_\nu/p}, \quad (102)$$

where  $\delta_3$  is as in (71). If  $\lambda_\alpha = 0$ , then

$$\begin{aligned} & \int_a^x q(\omega) \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_2} f_{j+1})(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_j)(\omega)|^{\lambda_\nu} \right. \right. \right. \\ & \quad \left. \left. \left. + |(D_{*a}^{\gamma_2} f_j)(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_{j+1})(\omega)|^{\lambda_\nu} \right] \right\} \right. \\ & \quad \left. + \left[ |(D_{*a}^{\gamma_2} f_M)(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_1)(\omega)|^{\lambda_\nu} \right. \right. \\ & \quad \left. \left. + |(D_{*a}^{\gamma_2} f_1)(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_M)(\omega)|^{\lambda_\nu} \right] \right\} d\omega \geq \\ & M^{1-(\frac{\lambda_\nu+\lambda_\beta}{p})} 2^{\left(\frac{\lambda_\nu+\lambda_\beta}{p}\right)} \varphi_2(x) \cdot \left\{ \int_a^x p(\omega) \right. \\ & \quad \left. \left[ \sum_{j=1}^M |(D_{*a}^\nu f_j)(\omega)|^p \right] d\omega \right\}^{\left(\frac{\lambda_\nu+\lambda_\beta}{p}\right)}, \end{aligned} \quad (103)$$

all  $a \leq x \leq b$ .

**Proof.** As Theorem 12 of [7], and Theorem 12 of [8].  $\square$

We give

**Theorem 56.** Let  $\nu \geq 2$  and  $\gamma_1 \geq 0$  such that  $2 \leq \nu - \gamma_1 < \frac{1}{p}$ ,  $0 < p < 1$ ,  $n := \lceil \nu \rceil$ . Consider  $f_i \in AC^n([a, b])$ ,  $i = 1, \dots, M \in \mathbb{N}$ , such that  $f_i^{(l)}(a) = 0$ ,  $l = 0, 1, \dots, n - 1$ ;  $i = 1, \dots, M$ ;  $a \leq x \leq b$ . Assume that  $D_{*a}^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, \dots, M$ ; each of which is of fixed sign a.e. on  $[a, b]$ . Consider also  $p(t) > 0$  and  $q(t) \geq 0$ , with all  $p(t)$ ,  $\frac{1}{p(t)}$ ,  $q(t)$ ,  $\frac{1}{q(t)} \in L_\infty(a, b)$ . Let  $\lambda_\alpha \geq \lambda_{\alpha+1} > 1$ .

Here  $\theta_3$  is as in (73),  $L(x)$  as in (50),  $P_1(x)$  as in (51),  $T(x)$  as in (52),  $\omega_1$  as in (53), and  $\Phi$  as in (54).

$$\begin{aligned} \int_a^x q(\omega) \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_1} f_j)(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_{j+1}(\omega)|^{\lambda_{\alpha+1}} \right. \right. \right. \\ \left. \left. \left. + |(D_{*a}^{\gamma_1} f_{j+1})(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_j(\omega)|^{\lambda_{\alpha+1}} \right] \right\} \right. \\ \left. + \left[ |(D_{*a}^{\gamma_1} f_1)(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_M(\omega)|^{\lambda_{\alpha+1}} \right. \right. \\ \left. \left. + |(D_{*a}^{\gamma_1} f_M)(\omega)|^{\lambda_\alpha} |D_{*a}^{\gamma_1+1} f_1(\omega)|^{\lambda_{\alpha+1}} \right] \right\} d\omega \geq \\ M^{1-\left(\frac{\lambda_\alpha+\lambda_{\alpha+1}}{p}\right)} 2^{\left(\frac{\lambda_\alpha+\lambda_{\alpha+1}}{p}\right)} \Phi(x) \cdot \\ \left[ \int_a^x p(\omega) \left( \sum_{j=1}^M |(D_{*a}^\nu f_j)(\omega)|^p \right) d\omega \right]^{\left(\frac{\lambda_\alpha+\lambda_{\alpha+1}}{p}\right)}, \end{aligned} \quad (104)$$

all  $a \leq x \leq b$ .

**Proof.** As Theorem 13 of [7] and Theorem 13 of [8].  $\square$

We have the special cases.

**Corollary 57.** (to Theorem 54,  $\lambda_\beta = 0$ ,  $p(t) = q(t) = 1$ ).

Then

$$\begin{aligned} \int_a^x \left( \sum_{j=1}^M |D_{*a}^{\gamma_1} f_j(\omega)|^{\lambda_\alpha} |D_{*a}^\nu f_j(\omega)|^{\lambda_\nu} \right) d\omega \\ \geq \delta_1^* \varphi_1(x) \left[ \int_a^x \left[ \sum_{j=1}^M |D_{*a}^\nu f_j(\omega)|^p \right] d\omega \right]^{\left(\frac{\lambda_\alpha+\lambda_\nu}{p}\right)}, \end{aligned} \quad (105)$$

all  $a \leq x \leq b$ .

In (105),  $(A_0(x)|_{\lambda_\beta=0})$  of  $\varphi_1(x)$  is given by (59).

**Corollary 58.** (to Theorem 55,  $\lambda_\alpha = 0$ ,  $p(t) = q(t) = 1$ ). It holds

$$\int_a^x \left\{ \left\{ \sum_{j=1}^{M-1} \left[ |(D_{*a}^{\gamma_2} f_{j+1})(\omega)|^{\lambda_\beta} |(D_{*a}^\nu f_j)(\omega)|^{\lambda_\nu} \right. \right. \right.$$

$$\begin{aligned}
& + |(D_{*a}^{\gamma_2} f_j)(\omega)|^{\lambda_\beta} |(D_{*a}^{\nu} f_{j+1})(\omega)|^{\lambda_\nu} \Big] \Big\} \\
& + \Big[ |(D_{*a}^{\gamma_2} f_M)(\omega)|^{\lambda_\beta} |(D_{*a}^{\nu} f_1)(\omega)|^{\lambda_\nu} \\
& + |(D_{*a}^{\gamma_2} f_1)(\omega)|^{\lambda_\beta} |(D_{*a}^{\nu} f_M)(\omega)|^{\lambda_\nu} \Big] d\omega \geq \\
& \left( M^{1 - \left( \frac{\lambda_\nu + \lambda_\beta}{p} \right)} \right) 2^{\left( \frac{\lambda_\nu + \lambda_\beta}{p} \right)} \varphi_2(x) \cdot \\
& \left\{ \int_a^x \left[ \sum_{j=1}^M |(D_{*a}^{\nu} f_j)(\omega)|^p \right] d\omega \right\}^{\left( \frac{\lambda_\nu + \lambda_\beta}{p} \right)}, \tag{106}
\end{aligned}$$

all  $a \leq x \leq b$ .

In (106),  $(A_0(x)|_{\lambda_\alpha=0})$  of  $\varphi_2(x)$  is given by (64).

Next we apply above results on the spherical shell.

## 4 Background-II

Here initially we follow [24], pp. 149-150 and [25], pp. 87-88. Let us denote by  $dx \equiv \lambda_{\mathbb{R}^N}(dx)$ ,  $N \in \mathbb{N}$ , the Lebesgue measure on  $\mathbb{R}^N$ , and  $S^{N-1} := \{x \in \mathbb{R}^N : |x| = 1\}$  the unit sphere on  $\mathbb{R}^N$ , where  $|\cdot|$  stands for the Euclidean norm in  $\mathbb{R}^N$ . Also denote the ball

$$B(0, R) := \{x \in \mathbb{R}^N : |x| < R\} \subseteq \mathbb{R}^N, \quad R > 0,$$

and the spherical shell

$$A := B(0, R_2) - \overline{B(0, R_1)}, \quad 0 < R_1 < R_2.$$

For  $x \in \mathbb{R}^N - \{0\}$  we can write uniquely  $x = r\omega$ , where  $r = |x| > 0$ , and  $\omega = \frac{x}{r} \in S^{N-1}$ ,  $|\omega| = 1$ . Clearly here

$$\mathbb{R}^N - \{0\} = (0, \infty) \times S^{N-1},$$

and the map

$$\Phi : \mathbb{R}^N - \{0\} \rightarrow S^{N-1} : \Phi(x) = \frac{x}{|x|}$$

is continuous.

Also  $\overline{A} = [R_1, R_2] \times S^{N-1}$ . Let us denote by  $d\omega \equiv \lambda_{S^{N-1}}(\omega)$  the surface measure on  $S^{N-1}$  to be defined as the image under  $\Phi$  of  $N \cdot \lambda_{\mathbb{R}^N}$  restricted to the Borel class of  $B(0, 1) - \{0\}$ . More precisely the last definition has as follows: let  $A \subset S^{N-1}$  be a Borel set, and let

$$\tilde{A} := \{ru : 0 < r < 1, u \in A\} \subset \mathbb{R}^N,$$

we define

$$\lambda_{S^{N-1}}(A) = N \cdot \lambda_{\mathbb{R}^N}(\tilde{A}).$$

$B_X$  here stands for the Borel class on space  $X$ .

We denote by

$$\omega_N \equiv \lambda_{S^{N-1}}(S^{N-1}) = \int_{S^{N-1}} d\omega = \frac{2\pi^{N/2}}{\Gamma(N/2)}$$

the surface area of  $S^{N-1}$  and we get the volume

$$|B(0, r)| = \frac{\omega_N r^N}{N} = \frac{2\pi^{N/2} r^N}{N\Gamma(N/2)},$$

so that

$$|B(0, 1)| = \frac{2\pi^{N/2}}{N\Gamma(N/2)}.$$

Clearly here

$$\text{Vol}(A) = |A| = \frac{\omega_N (R_2^N - R_1^N)}{N} = \frac{2\pi^{N/2} (R_2^N - R_1^N)}{N\Gamma(N/2)}.$$

Next, define

$$\psi : (0, \infty) \times S^{N-1} \rightarrow \mathbb{R}^N - \{0\}$$

by  $\psi(r, \omega) := r\omega$ ,  $\psi$  is one to one and onto function, thus

$$(r, \omega) \equiv \psi^{-1}(x) = (|x|, \Phi(x))$$

are called the polar coordinates of  $x \in \mathbb{R}^N - \{0\}$ .

Finally, define the measure  $R_N$  on  $((0, \infty), \mathcal{B}_{(0, \infty)})$  by

$$R_N(\Gamma) = \int_{\Gamma} r^{N-1} dr, \text{ any } \Gamma \in \mathcal{B}_{(0, \infty)}.$$

We mention the very important theorem

**Theorem 59.** (see exercise 6, pp. 149-150 in [24] and Theorem 5.2.2 pp. 87-88 of [25]) *We have that  $\lambda_{\mathbb{R}^N} = (R_N \times \lambda_{S^{N-1}}) \circ \psi^{-1}$  on  $\mathcal{B}_{\mathbb{R}^N - \{0\}}$ .*

*In particular, if  $f$  is a non-negative Borel measurable function on  $(\mathbb{R}^N, \mathcal{B}_{\mathbb{R}^N})$ , then the Lebesgue integral*

$$\begin{aligned} \int_{\mathbb{R}^N} f(x) dx &= \int_{(0, \infty)} r^{N-1} \left( \int_{S^{N-1}} f(r\omega) \lambda_{S^{N-1}}(d\omega) \right) dr \\ &= \int_{S^{N-1}} \left( \int_{(0, \infty)} f(r\omega) r^{N-1} dr \right) \lambda_{S^{N-1}}(d\omega). \end{aligned} \quad (107)$$



Clearly (107) is true for  $f$  a Borel integrable function taking values in  $\mathbb{R}$ .

Based on Theorem 59 in [5] we proved the next result which is the main tool of this section.

**Proposition 60.** Let

$$f : A \rightarrow \mathbb{R},$$

be a Lebesgue integrable function, where

$$A := B(0, R_2) - \overline{B(0, R_1)}, \quad 0 < R_1 < R_2. \quad (108)$$

Then

$$\int_A f(x) dx = \int_{S^{N-1}} \left( \int_{R_1}^{R_2} f(r\omega) r^{N-1} dr \right) d\omega.$$

We make

**Remark 61.** Let  $F : \overline{A} = [R_1, R_2] \times S^{N-1} \rightarrow \mathbb{R}$  and for each  $\omega \in S^{N-1}$  define

$$g_\omega(r) := F(r\omega) = F(x),$$

where  $x \in \overline{A}$ , with  $A := B(0, R_2) - \overline{B(0, R_1)}$ ;  $0 < R_1 \leq r \leq R_2$ ,  $r = |x|$ ,  $\omega = \frac{x}{r} \in S^{N-1}$ .

For each  $\omega \in S^{N-1}$  we assume that  $g_\omega \in AC^n([R_1, R_2])$ , where  $n := \lceil \nu \rceil$ ,  $\nu \geq 0$ .

Thus, by Corollary 8, for almost all  $r \in [R_1, R_2]$ , there exists the Caputo fractional derivative

$$D_{*R_1}^\nu g_\omega(r) = \frac{1}{\Gamma(n-\nu)} \int_{R_1}^r (r-t)^{n-\nu-1} g_\omega^{(n)}(t) dt, \quad (109)$$

for all  $\omega \in S^{N-1}$ .

Now we are ready to give

**Definition 62.** Let  $F : \overline{A} \rightarrow \mathbb{R}$ ,  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$  such that  $F(\cdot\omega) \in AC^n([R_1, R_2])$ , for all  $\omega \in S^{N-1}$ .

We call the Caputo radial fractional derivative the following function

$$\frac{\partial_{*R_1}^\nu F(x)}{\partial r^\nu} := \frac{1}{\Gamma(n-\nu)} \int_{R_1}^r (r-t)^{n-\nu-1} \frac{\partial^n F(t\omega)}{\partial r^n} dt, \quad (110)$$

where  $x \in \overline{A}$ , i.e.  $x = r\omega$ ,  $r \in [R_1, R_2]$ ,  $\omega \in S^{N-1}$ .

Clearly

$$\begin{aligned} \frac{\partial_{*R_1}^0 F(x)}{\partial r^0} &= F(x), \\ \frac{\partial_{*R_1}^\nu F(x)}{\partial r^\nu} &= \frac{\partial^\nu F(x)}{\partial r^\nu}, \text{ if } \nu \in \mathbb{N}. \end{aligned}$$

Above function (110) exists almost every where for  $x \in \overline{A}$ .

We justify this next.

**Note 63.** Call

$$\Lambda_1 := \left\{ r \in [R_1, R_2] : \frac{\partial_{*R_1}^\nu F(x)}{\partial r^\nu} \text{ does not exist} \right\}. \quad (111)$$

We have that Lebesgue measure  $\lambda_{\mathbb{R}}(\Lambda_1) = 0$ .

Call  $\Lambda_N := \Lambda_1 \times S^{N-1}$ . So there exists a Borel set  $\Lambda_1^* \subset [R_1, R_2]$ , such that  $\Lambda_1 \subset \Lambda_1^*$ ,  $\lambda_{\mathbb{R}}(\Lambda_1^*) = \lambda_{\mathbb{R}}(\Lambda_1) = 0$ , thus  $\lambda_N(\Lambda_1^*) = 0$ .

Consider now  $\Lambda_N^* := \Lambda_1^* \times S^{N-1} \subset \bar{A}$ , which is a Borel set of  $\mathbb{R}^N - \{0\}$ . Clearly then by Theorem 59,  $\lambda_{\mathbb{R}^N}(\Lambda_N^*) = 0$ , but  $\Lambda_N \subset \Lambda_N^*$ , implying  $\lambda_{\mathbb{R}^N}(\Lambda_N) = 0$ .

Consequently (110) exists a.e. in  $x$  with respect to  $\lambda_{\mathbb{R}^N}$  on  $\bar{A}$ .

We give the following fundamental representation result.

**Theorem 64.** *Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $F : \bar{A} \rightarrow \mathbb{R}$  with  $F \in L_1(A)$ . Assume that  $F(\cdot\omega) \in AC^n([R_1, R_2])$  for all  $\omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu F(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$  for all  $\omega \in S^{N-1}$ .*

*Further assume that  $\frac{\partial_{*R_1}^\nu F(x)}{\partial r^\nu} \in L_\infty(A)$ . More precisely, for these  $r \in [R_1, R_2]$ , for each  $\omega \in S^{N-1}$ , for which  $D_{*R_1}^\nu F(r\omega)$  takes real values, there exists  $M_1 > 0$  such that  $|D_{*R_1}^\nu F(r\omega)| \leq M_1$ .*

*We suppose that  $\frac{\partial^i F(R_1\omega)}{\partial r^j} = 0$ ,  $j = 0, 1, \dots, n-1$ , for every  $\omega \in S^{N-1}$ . Then*

$$\begin{aligned} \frac{\partial_{*R_1}^\gamma F(x)}{\partial r^\gamma} &= D_{*R_1}^\gamma F(r\omega) = \\ &= \frac{1}{\Gamma(\nu - \gamma)} \int_{R_1}^r (r-t)^{\nu-\gamma-1} (D_{*R_1}^\nu F)(t\omega) dt, \end{aligned} \quad (112)$$

*true  $\forall x \in \bar{A}$ , i.e. true  $\forall r \in [R_1, R_2]$  and  $\forall \omega \in S^{N-1}$ ,  $\gamma > 0$ .*

*Here*

$$D_{*R_1}^\gamma F(\cdot\omega) \in AC([R_1, R_2]), \quad (113)$$

*$\forall \omega \in S^{N-1}$ ,  $\gamma > 0$ .*

*Furthermore*

$$\frac{\partial_{*R_1}^\gamma F(x)}{\partial r^\gamma} \in L_\infty(A), \quad \gamma > 0. \quad (114)$$

*In particular, it holds*

$$F(x) = F(r\omega) = \frac{1}{\Gamma(\nu)} \int_{R_1}^r (r-t)^{\nu-1} (D_{*R_1}^\nu F)(t\omega) dt, \quad (115)$$

*true  $\forall x \in \bar{A}$ , i.e. true  $\forall r \in [R_1, R_2]$  and  $\forall \omega \in S^{N-1}$ , and*

$$F(\cdot\omega) \in AC([R_1, R_2]), \quad \forall \omega \in S^{N-1}. \quad (116)$$

**Proof.** By our assumptions and Theorem 16, Corollary 14, we have valid (112) and (115). Also (113) is clear, see [5]. Property (116) is easy to prove.

Fixing  $r \in [R_1, R_2]$ , the function

$$\delta_r(t, \omega) := (r-t)^{\nu-\gamma-1} D_{*R_1}^\nu F(t\omega)$$

is measurable on

$$\left([R_1, r] \times S^{N-1}, \overline{\mathcal{B}_{[R_1, r]} \times \mathcal{B}_{S^{N-1}}}\right).$$

Here  $\overline{\mathcal{B}_{[R_1, r]} \times \mathcal{B}_{S^{N-1}}}$  stands for the complete  $\sigma$ -algebra generated by  $\mathcal{B}_{[R_1, r]} \times \mathcal{B}_{S^{N-1}}$ , where  $\mathcal{B}_X$  stands for the completion of  $\mathcal{B}_X$ .

Then we get that

$$\int_{S^{N-1}} \left( \int_{R_1}^r |\delta_r(t, \omega)| dt \right) d\omega = \int_{S^{N-1}} \left( \int_{R_1}^r (r-t)^{\nu-\gamma-1} |D_{*R_1}^\nu F(t\omega)| dt \right) d\omega \leq \quad (117)$$

$$\left\| \frac{\partial_{*R_1}^\nu F(x)}{\partial r^\nu} \right\|_{\infty, ([R_1, r] \times S^{N-1})} \left( \int_{S^{N-1}} \left( \int_{R_1}^r (r-t)^{\nu-\gamma-1} dt \right) d\omega \right) \quad (118)$$

$$= \left\| \frac{\partial_{*R_1}^\nu F(x)}{\partial r^\nu} \right\|_{\infty, ([R_1, r] \times S^{N-1})} \left( \frac{2\pi^{N/2}}{\Gamma(N/2)} \right) \frac{(r-R_1)^{\nu-\gamma}}{(\nu-\gamma)} \leq \left\| \frac{\partial_{*R_1}^\nu F(x)}{\partial r^\nu} \right\|_{\infty, A} \left( \frac{2\pi^{N/2}}{\Gamma(N/2)} \right) \frac{(R_2-R_1)^{\nu-\gamma}}{(\nu-\gamma)} < \infty. \quad (119)$$

Hence  $\delta_r(t, \omega)$  is integrable on

$$\left([R_1, r] \times S^{N-1}, \overline{\mathcal{B}_{[R_1, r]} \times \mathcal{B}_{S^{N-1}}}\right).$$

Consequently, by Fubini's theorem and (112), we obtain that  $D_{*R_1}^\gamma F(r\omega)$ ,  $\nu \geq \gamma + 1$ ,  $\gamma > 0$  is integrable in  $\omega$  over  $(S^{N-1}, \overline{\mathcal{B}_{S^{N-1}}})$ . So we have that  $D_{*R_1}^\gamma F(r\omega)$  is continuous in  $r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ , and measurable in  $\omega \in S^{N-1}$ ,  $\forall r \in [R_1, R_2]$ . So, it is a Carathéodory function. Here  $[R_1, R_2]$  is a separable metric space and  $S^{N-1}$  is a measurable space, and the function takes values in  $\mathbb{R}^* := \mathbb{R} \cup \{\pm\infty\}$ , which is a metric space. Therefore by Theorem 20.15, p. 156 of [1],  $(D_{*R_1}^\gamma F)(r\omega)$  is jointly  $(\mathcal{B}_{[R_1, R_2]} \times \overline{\mathcal{B}_{S^{N-1}}})$ -measurable on  $[R_1, R_2] \times S^{N-1} = A$ , that is Lebesgue measurable on  $A$ . Indeed then we have that

$$|D_{*R_1}^\gamma F(r\omega)| \leq \frac{1}{\Gamma(\nu-\gamma)}$$

$$\int_{R_1}^r (r-t)^{\nu-\gamma-1} |D_{*R_1}^\nu F(t\omega)| dt \quad (120)$$

$$\leq \frac{\|D_{*R_1}^\nu F(\cdot\omega)\|_{\infty, [R_1, R_2]}}{\Gamma(\nu-\gamma)} \left( \int_{R_1}^r (r-t)^{\nu-\gamma-1} dt \right) \leq \quad (121)$$

$$\frac{M_1}{\Gamma(\nu-\gamma)} \frac{(r-R_1)^{\nu-\gamma}}{(\nu-\gamma)} \leq \frac{M_1}{\Gamma(\nu-\gamma-1)} (R_2-R_1)^{\nu-\gamma} := \tau < \infty,$$

for all  $\omega \in S^{N-1}$  and for all  $r \in [R_1, R_2]$ .

I.e. we proved that

$$\left| D_{*R_1}^\gamma F(r\omega) \right| \leq \tau < \infty, \forall \omega \in S^{N-1}, \text{ and } \forall r \in [R_1, R_2]. \quad (122)$$

Hence proving  $\frac{\partial_{*R_1}^\gamma F(x)}{\partial r^\gamma} \in L_\infty(A)$ ,  $\gamma > 0$ . We have completed our proof.

□

## 5 Main results on a spherical shell

### 5.1 Results involving one function

We give

**Theorem 65.** *Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, \dots, l \in \mathbb{N}$ ,  $n := \lceil \nu \rceil$ , and  $0 \leq \gamma_1 < \gamma_2 \leq \gamma_3 \leq \dots \leq \gamma_l$ . Here  $f : A \rightarrow \mathbb{R}$  is as in Theorem 64. Let  $r_i > 0 : \sum_{i=1}^l r_i = p$ . If  $\gamma_1 = 0$  we set  $r_1 = 1$ . Let  $s_1, s'_1 > 1 : \frac{1}{s_1} + \frac{1}{s'_1} = 1$ , and  $s_2, s'_2 > 1 : \frac{1}{s_2} + \frac{1}{s'_2} = 1$ , and  $p > s_2$ ,  $N \geq 2$ . Denote*

$$Q_1(R_2) := \left( \frac{R_2^{(N-1)s'_1+1} - R_1^{(N-1)s'_1+1}}{(N-1)s'_1+1} \right)^{1/s'_1}, \quad (123)$$

$$Q_2(R_2) := \left( \frac{R_2^{(1-N)\frac{s'_2}{p}+1} - R_1^{(1-N)\frac{s'_2}{p}+1}}{(1-N)\frac{s'_2}{p}+1} \right)^{p/s'_2}, \quad (124)$$

and

$$\sigma := \frac{p - s_2}{ps_2}. \quad (125)$$

Also call

$$C := Q_1(R_2) Q_2(R_2) \prod_{i=1}^l \left\{ \frac{\sigma^{r_i \sigma}}{(\Gamma(\nu - \gamma_i))^{r_i} (\nu - \gamma_i - 1 + \sigma)^{r_i \sigma}} \right\} \frac{(R_2 - R_1)^{\left( \sum_{i=1}^l (\nu - \gamma_i - 1) r_i + \frac{p}{s_2} + \frac{1}{s_1} - 1 \right)}}{\left( \left( \sum_{i=1}^l (\nu - \gamma_i - 1) r_i s_1 \right) + s_1 \left( \frac{p}{s_2} - 1 \right) + 1 \right)^{1/s_1}}. \quad (126)$$

Then

$$\begin{aligned} \int_A \prod_{i=1}^l \left| \frac{\partial_{*R_1}^{\gamma_i} f(x)}{\partial r^{\gamma_i}} \right|^{r_i} dx &\leq \\ C \int_A \left| \frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \right|^p dx. \end{aligned} \quad (127)$$

**Proof.** Clearly here  $f(\cdot\omega)$  fulfills all the assumptions of Theorem 24,  $\forall\omega \in S^{N-1}$ .

We set there  $q_1(r) = q_2(r) := r^{N-1}$ ,  $r \in [R_1, R_2]$ .

Hence by (30) we have

$$\begin{aligned} & \int_{R_1}^{R_2} r^{N-1} \prod_{i=1}^l \left| \frac{\partial_{*R_1}^{\gamma_i} f(r\omega)}{\partial r^{\gamma_i}} \right|^{r_i} dr \leq \\ & C \int_{R_1}^{R_2} r^{N-1} \left| \frac{\partial_{*R_1}^{\nu} f(r\omega)}{\partial r^{\nu}} \right|^p dr, \forall \omega \in S^{N-1}. \end{aligned} \quad (128)$$

Therefore it holds

$$\begin{aligned} & \int_{S^{N-1}} \left( \int_{R_1}^{R_2} r^{N-1} \prod_{i=1}^l \left| \frac{\partial_{*R_1}^{\gamma_i} f(r\omega)}{\partial r^{\gamma_i}} \right|^{r_i} dr \right) d\omega \leq \\ & C \int_{S^{N-1}} \left( \int_{R_1}^{R_2} r^{N-1} \left| \frac{\partial_{*R_1}^{\nu} f(r\omega)}{\partial r^{\nu}} \right|^p dr \right) d\omega. \end{aligned} \quad (129)$$

Using conclusion of Theorem 64 and Proposition 60 we derive (127).  $\square$

We continue with the following extreme case.

**Theorem 66.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, \dots, l \in \mathbb{N}$ ,  $n := \lceil \nu \rceil$ , and  $0 \leq \gamma_1 < \gamma_2 \leq \gamma_3 \leq \dots \leq \gamma_l$ . Here  $f : \bar{A} \rightarrow \mathbb{R}$  as in Theorem 64. Let  $r_i > 0 : \sum_{i=1}^l r_i = r$ . If  $\gamma_1 = 0$  we set  $r_1 = 1$ ,  $N \geq 2$ .  
Call

$$\widetilde{M} := \frac{R_2^{N-1} (R_2 - R_1)^{r\nu - \sum_{i=1}^l r_i \gamma_i + 1}}{\prod_{i=1}^l (\Gamma(\nu - \gamma_i + 1))^{r_i} (r\nu - \sum_{i=1}^l r_i \gamma_i + 1)} > 0. \quad (130)$$

Then

$$\begin{aligned} & \int_A \left( \prod_{i=1}^l \left| \frac{\partial_{*R_1}^{\gamma_i} f(x)}{\partial r^{\gamma_i}} \right|^{r_i} \right) dx \leq \\ & \widetilde{M} M_1^r \frac{2\pi^{N/2}}{\Gamma(N/2)}. \end{aligned} \quad (131)$$

**Proof.** Clearly here  $f(\cdot\omega)$  fulfills all the assumptions of Theorem 26,  $\forall\omega \in S^{N-1}$ .

We set  $\tilde{q}(r) = r^{N-1}$ ,  $r \in [R_1, R_2]$ .

Hence by (35) we have

$$\int_{R_1}^{R_2} r^{N-1} \prod_{i=1}^l |D_{*R_1}^{\gamma_i} f(r\omega)|^{r_i} dr \leq$$

$$\left( \frac{R_2^{N-1} (R_2 - R_1)^{r\nu - \sum_{i=1}^l r_i \gamma_i + 1}}{\prod_{i=1}^l (\Gamma(\nu - \gamma_i + 1))^{r_i} (r\nu - \sum_{i=1}^l r_i \gamma_i + 1)} \right) \left\| \frac{\partial_{*R_1}^\nu f(\cdot\omega)}{\partial r^\nu} \right\|_{\infty, [R_1, R_2]} \leq \widetilde{M} M_1^r, \quad \forall \omega \in S^{N-1}. \quad (132)$$

Hence

$$\int_{S^{N-1}} \left( \int_{R_1}^{R_2} r^{N-1} \prod_{i=1}^l |D_{*R_1}^{\gamma_i} f(r\omega)|^{r_i} dr \right) d\omega \leq \widetilde{M} M_1^r \frac{2\pi^{N/2}}{\Gamma(N/2)}. \quad (133)$$

Using the conclusion of Theorem 64 and Proposition 60 we derive (131).  $\square$

## 5.2 Results involving two function

We give

We need to make

**Assumption 67.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, 2$ ,  $n := \lceil \nu \rceil$ ,  $f_1, f_2 : \bar{A} \rightarrow \mathbb{R}$  with  $f_1, f_2 \in L_1(A)$ , where  $A := B(0, R_2) - \bar{B}(0, R_1)$ ,  $0 < R_1 < R_2$ . Assume that  $f_1(\cdot\omega), f_2(\cdot\omega) \in AC^n([R_1, R_2])$  for all  $\omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f_i(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ , for all  $\omega \in S^{N-1}$ ;  $i = 1, 2$ . Further assume that  $\frac{\partial_{*R_1}^\nu f_i(x)}{\partial r^\nu} \in L_\infty(A)$ ,  $i = 1, 2$ . More precisely, for these  $r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ , for which  $D_{*R_1}^\nu f_i(r\omega)$  takes real values, there exists  $M_i > 0$  such that

$$|D_{*R_1}^\nu f_i(r\omega)| \leq M_i, \quad \text{for } i = 1, 2. \quad (134)$$

We suppose that

$$\frac{\partial^j f_i(R_1\omega)}{\partial r^j} = 0, \quad j = 0, 1, \dots, n-1,$$

$\forall \omega \in S^{N-1}$ ;  $i = 1, 2$ .

Let  $\lambda_\nu > 0$ , and  $\lambda_\alpha, \lambda_\beta \geq 0$ , such that  $\lambda_\nu < p$ , where  $p > 1$ . If  $\gamma_1 = 0$  we set  $\lambda_\alpha = 1$  and if  $\gamma_2 = 0$  we set  $\lambda_\beta = 1$ , here  $N \geq 2$ .

**Assumption 67\*.** (continuation of Assumption 67)

Set

$$P_k(w) := \int_{R_1}^w (w-t)^{(\nu-\gamma_k-1)p/(p-1)} t^{\left(\frac{1-N}{p-1}\right)} dt, \quad (135)$$

$k = 1, 2$ ,  $R_1 \leq w \leq R_2$ ,

$$A(w) := \frac{w^{(N-1)(1-\frac{\lambda_\nu}{p})} (P_1(w))^{\lambda_\alpha(\frac{p-1}{p})} (P_2(w))^{\lambda_\beta(\frac{p-1}{p})}}{(\Gamma(\nu-\gamma_1))^{\lambda_\alpha} (\Gamma(\nu-\gamma_2))^{\lambda_\beta}}, \quad (136)$$

$$A_0(R_2) := \left( \int_{R_1}^{R_2} (A(w))^{p/(p-\lambda_\nu)} dw \right)^{\frac{(p-\lambda_\nu)}{p}}. \quad (137)$$

We present

**Theorem 68.** *All here as in Assumption 67,  $67^*$ , especially assume  $\lambda_\alpha > 0$ ,  $\lambda_\beta = 0$  and  $p = \lambda_\alpha + \lambda_\nu > 1$ . Then*

$$\begin{aligned} & \int_A \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \\ & \left. \left| \frac{\partial_{*R_1}^{\gamma_1} f_2(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] dx \leq \\ & (A_0(R_2)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{p} \right)^{\left( \frac{\lambda_\nu}{p} \right)} \\ & \int_A \left[ \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^p + \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^p \right] dx. \end{aligned} \quad (138)$$

**Proof.** We apply here Theorem 30 for every  $\omega \in S^{N-1}$ , here  $p(r) = q(r) = r^{N-1}$ ,  $r \in [R_1, R_2]$ . Use of Theorem 64 and Proposition 60. So proof is similar to the proof of Theorem 65.  $\square$

It follows the counterpart of the last theorem.

**Theorem 69.** *All here as in Assumption 67,  $67^*$ , especially suppose  $\lambda_\alpha = 0$ ,  $\lambda_\beta > 0$ ,  $p = \lambda_\nu + \lambda_\beta > 1$ .*

Denote

$$\delta_3 := \begin{cases} 2^{\lambda_\beta/\lambda_\nu} - 1, & \text{if } \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\beta \leq \lambda_\nu. \end{cases} \quad (139)$$

Then

$$\begin{aligned} & \int_A \left[ \left| \frac{\partial_{*R_1}^{\gamma_2} f_2(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \\ & \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_1(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] dx \leq \\ & (A_0(R_2)|_{\lambda_\alpha=0}) 2^{\lambda_\beta/p} \left( \frac{\lambda_\nu}{p} \right)^{\left( \frac{\lambda_\nu}{p} \right)} \delta_3^{\lambda_\nu/p} \\ & \int_A \left[ \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^p + \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^p \right] dx. \end{aligned} \quad (140)$$

**Proof.** Based on Theorem 31, similar to the proof of Theorem 68.  $\square$

**Theorem 70.** All here as in Assumption 67, 67\*, especially suppose  $\lambda_\nu$ ,  $\lambda_\alpha, \lambda_\beta > 0$ ,  $p = \lambda_\alpha + \lambda_\beta + \lambda_\nu > 1$ .

Denote

$$\tilde{\gamma}_1 := \begin{cases} 2^{\left(\frac{\lambda_\alpha + \lambda_\beta}{\lambda_\nu}\right)} - 1, & \text{if } \lambda_\alpha + \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\alpha + \lambda_\beta \leq \lambda_\nu. \end{cases} \quad (141)$$

Then

$$\begin{aligned} & \int_A \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_2} f_2(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \\ & \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_1(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^{\gamma_1} f_2(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] dx \leq \\ & A_0(R_2) \left( \frac{\lambda_\nu}{(\lambda_\alpha + \lambda_\beta)p} \right)^{(\lambda_\nu/p)} \left[ \lambda_\alpha^{\lambda_\nu/p} + 2^{(\lambda_\alpha + \lambda_\beta)/p} (\tilde{\gamma}_1 \lambda_\beta)^{\lambda_\nu/p} \right] \\ & \int_A \left( \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^p + \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^p \right) dx. \end{aligned} \quad (142)$$

**Proof.** Based on Theorem 32, similar to the proof of Theorem 68.  $\square$

We give the next special important case

**Theorem 71.** All as in Assumption 67 without  $\lambda_\nu$  there. Here  $\gamma_2 = \gamma_1 + 1$ ,  $\lambda_\alpha \geq 0$ ,  $\lambda_\beta := \lambda_{\alpha+1} \in (0, 1)$ , and  $p = \lambda_\alpha + \lambda_{\alpha+1} > 1$ .

Denote

$$\theta_3 := \begin{cases} 2^{(\lambda_\alpha/\lambda_{\alpha+1})} - 1 & \text{if } \lambda_\alpha \geq \lambda_{\alpha+1} \\ 1, & \text{if } \lambda_\alpha \leq \lambda_{\alpha+1}, \end{cases} \quad (143)$$

$$\begin{aligned} L(R_2) &:= \left[ 2^{\frac{(1 - \lambda_{\alpha+1})}{(N - \lambda_{\alpha+1})}} \right. \\ & \left. \left( R_2^{\frac{N - \lambda_{\alpha+1}}{1 - \lambda_{\alpha+1}}} - R_1^{\frac{N - \lambda_{\alpha+1}}{1 - \lambda_{\alpha+1}}} \right) \right]^{(1 - \lambda_{\alpha+1})} \left( \frac{\theta_3 \lambda_{\alpha+1}}{p} \right)^{\lambda_{\alpha+1}}, \end{aligned} \quad (144)$$

and

$$P_1(R_2) := \int_{R_1}^{R_2} (R_2 - t)^{(\nu - \gamma_1 - 1)p/(p-1)} t^{\left(\frac{1-N}{p-1}\right)} dt, \quad (145)$$

$$\Phi(R_2) := L(R_2) \left( \frac{P_1(R_2)^{(p-1)}}{(\Gamma(\nu - \gamma_1))^p} \right) 2^{p-1}. \quad (146)$$

Then

$$\begin{aligned} & \int_A \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_1+1} f_2(x)}{\partial r^{\gamma_1+1}} \right|^{\lambda_{\alpha+1}} + \right. \\ & \left. \left| \frac{\partial_{*R_1}^{\gamma_1} f_2(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_1+1} f_1(x)}{\partial r^{\gamma_1+1}} \right|^{\lambda_{\alpha+1}} \right] dx \end{aligned}$$



$$\leq \Phi(R_2) \int_A \left( \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^p + \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^p \right) dx. \quad (147)$$

**Proof.** Based on Theorem 33, similar to the proof of Theorem 68.  $\square$

We give an  $L_\infty$  result on the shell.

We need to make

**Assumption 72.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, 2$ ,  $n := \lfloor \nu \rfloor$ ,  $f_1, f_2 : \bar{A} \rightarrow \mathbb{R}$  with  $f_1, f_2 \in L_1(A)$ , where  $\mathbb{R}^N \supseteq A := B(0, R_2) - \bar{B}(0, R_1)$ ,  $0 < R_1 < R_2$ ,  $N \geq 2$ . Assume that  $f_1(\cdot\omega), f_2(\cdot\omega) \in AC^n([R_1, R_2])$  for all  $\omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f_i(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ , for all  $\omega \in S^{N-1}$ ;  $i = 1, 2$ . Further assume that  $\frac{\partial_{*R_1}^\nu f_i(x)}{\partial r^\nu} \in L_\infty(A)$ ,  $i = 1, 2$ . More precisely, for these  $r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ , for which  $D_{*R_1}^\nu f_i(r\omega)$  takes real values, there exists  $M_i > 0$  such that

$$|D_{*R_1}^\nu f_i(r\omega)| \leq M_i, \text{ for } i = 1, 2. \quad (148)$$

We suppose that

$$\frac{\partial^j f_i(R_1\omega)}{\partial r^j} = 0, j = 0, 1, \dots, n-1,$$

$\forall \omega \in S^{N-1}$ ;  $i = 1, 2$ .

Let  $\lambda_\nu, \lambda_\alpha, \lambda_\beta \geq 0$ . If  $\gamma_1 = 0$  we set  $\lambda_\alpha = 1$  and if  $\gamma_2 = 0$  we set  $\lambda_\beta = 1$ .

We present

**Theorem 73.** All as in Assumption 72.

Set

$$T(R_2 - R_1) := \frac{R_2^{N-1}}{(\nu\lambda_\alpha - \gamma_1\lambda_\alpha + \nu\lambda_\beta - \gamma_2\lambda_\beta + 1)} \cdot \frac{(R_2 - R_1)^{(\nu\lambda_\alpha - \gamma_1\lambda_\alpha + \nu\lambda_\beta - \gamma_2\lambda_\beta + 1)}}{(\Gamma(\nu - \gamma_1 + 1))^{\lambda_\alpha} (\Gamma(\nu - \gamma_2 + 1))^{\lambda_\beta}}. \quad (149)$$

Then

$$\begin{aligned} & \int_A \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_2} f_2(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \\ & \quad \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_1(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^{\gamma_1} f_2(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_2(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] dx \\ & \leq T(R_2 - R_1) \frac{\pi^{N/2}}{\Gamma(N/2)} \\ & \quad \left[ M_1^{2(\lambda_\alpha + \lambda_\nu)} + M_1^{2\lambda_\beta} + M_2^{2\lambda_\beta} + M_2^{2(\lambda_\alpha + \lambda_\nu)} \right]. \end{aligned} \quad (150)$$

**Proof.** Apply Theorem 38 for every  $\omega \in S^{N-1}$ , here  $p(r) = r^{N-1}$ ,  $r \in [R_1, R_2]$ . It goes as the proof of Theorem 66. Finally use Theorem 64 and Proposition 60.  $\square$

### 5.3 Results involving several function

We need to make

**Assumption 74.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, 2$ ,  $n := \lceil \nu \rceil$ ,  $f_j : \bar{A} \rightarrow \mathbb{R}$  with  $f_j \in L_1(A)$ ,  $j = 1, \dots, M$ ,  $M \in \mathbb{N}$ , where  $\mathbb{R}^N \supseteq A := B(0, R_2) - \overline{B(0, R_1)}$ ,  $0 < R_1 < R_2$ ,  $N \geq 2$ . Assume that  $f_j(\cdot\omega) \in AC^n([R_1, R_2])$  for all  $\omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f_j(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ , for all  $\omega \in S^{N-1}$ ;  $j = 1, \dots, M$ . Further assume that  $\frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \in L_\infty(A)$ ,  $j = 1, \dots, M$ . More precisely, for these  $r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ , for which  $D_{*R_1}^\nu f_j(r\omega)$  takes real values, there exists  $M_j > 0$  such that

$$|D_{*R_1}^\nu f_j(r\omega)| \leq M_j, \text{ for } j = 1, \dots, M. \quad (151)$$

We suppose that

$$\frac{\partial^j f_j(R_1\omega)}{\partial r^k} = 0, k = 0, 1, \dots, n-1,$$

$\forall \omega \in S^{N-1}$ ;  $j = 1, \dots, M$ .

Let  $\lambda_\nu > 0$ , and  $\lambda_\alpha, \lambda_\beta \geq 0$ , such that  $\lambda_\nu < p$ , where  $p > 1$ . If  $\gamma_1 = 0$  we set  $\lambda_\alpha = 1$  and if  $\gamma_2 = 0$  we set  $\lambda_\beta = 1$ .

We give

**Theorem 75.** Let  $f_j, j = 1, \dots, M$ , as in Assumption 74. Let  $\lambda_\nu > 0$ , and  $\lambda_\alpha > 0$ ;  $\lambda_\beta \geq 0$ ,  $p := \lambda_\alpha + \lambda_\nu > 1$ . Set

$$P_k(w) := \int_{R_1}^w (w-t)^{(\nu-\gamma_k-1)\frac{p}{(p-1)}} t^{\left(\frac{1-N}{p-1}\right)} dt, \quad (152)$$

$k = 1, 2$ ,  $R_1 \leq w \leq R_2$ ,

$$A(w) := \frac{w^{(N-1)(1-\frac{\lambda_\nu}{p})} (P_1(w))^{\lambda_\alpha(\frac{p-1}{p})} (P_2(w))^{\lambda_\beta(\frac{p-1}{p})}}{(\Gamma(\nu-\gamma_1))^{\lambda_\alpha} (\Gamma(\nu-\gamma_2))^{\lambda_\beta}}, \quad (153)$$

$$A_0(R_2) := \left( \int_{R_1}^{R_2} (A(w))^{p/\lambda_\alpha} dw \right)^{\lambda_\alpha/p}. \quad (154)$$

Take the case of  $\lambda_\beta = 0$ . Then

$$\begin{aligned} & \sum_{j=1}^M \int_A \left| \frac{\partial_{*R_1}^{\gamma_1} f_j(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^{\lambda_\nu} dx \\ & \leq (A_0(R_2)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{p} \right)^{\left(\frac{\lambda_\nu}{p}\right)} \\ & \left[ \sum_{j=1}^M \left( \int_A \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^p dx \right) \right]. \end{aligned} \quad (155)$$

**Proof.** As in Theorem 68, based on Theorem 45.  $\square$

We continue with

**Theorem 76.** *All basic assumptions as in Theorem 75. Let  $\lambda_\nu > 0$ ,  $\lambda_\alpha = 0$ ;  $\lambda_\beta > 0$ ,  $p := \lambda_\nu + \lambda_\beta > 1$ ,  $P_2$  defined by (152).*

Now it is

$$A(w) := \frac{w^{(N-1)(1-\frac{\lambda_\nu}{p})} (P_2(w))^{\lambda_\beta(\frac{p-1}{p})}}{(\Gamma(\nu - \gamma_2))^{\lambda_\beta}}, \quad (156)$$

$$A_0(R_2) := \left( \int_{R_1}^{R_2} (A(w))^{p/\lambda_\beta} dw \right)^{\lambda_\beta/p}. \quad (157)$$

Denote

$$\delta_3 := \begin{cases} 2^{\lambda_\beta/\lambda_\nu} - 1, & \text{if } \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\beta \leq \lambda_\nu. \end{cases} \quad (158)$$

Call

$$\varphi_2(R_2) := A_0(R_2) 2^{\lambda_\beta/p} \left( \frac{\lambda_\nu}{p} \right)^{\lambda_\nu/p} \delta_3^{\lambda_\nu/p}. \quad (159)$$

Then

$$\begin{aligned} \int_A \left\{ \left\{ \sum_{j=1}^{M-1} \left[ \left| \frac{\partial_{*R_1}^{\gamma_2} f_{j+1}(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \right. \right. \\ \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_j(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_{j+1}(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] \right\} + \\ \left[ \left| \frac{\partial_{*R_1}^{\gamma_2} f_M(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \\ \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_1(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_M(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] \Big\} dx \leq \\ 2\varphi_2(R_2) \left[ \sum_{j=1}^M \left( \int_A \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^p dx \right) \right]. \end{aligned} \quad (160)$$

**Proof.** As in Theorem 68, based on Theorem 46.  $\square$

We present the general case

**Theorem 77.** *All basic assumptions as in Theorem 75. Here  $\lambda_\nu, \lambda_\alpha, \lambda_\beta > 0$ ,  $p := \lambda_\alpha + \lambda_\beta + \lambda_\nu > 1$ ,  $P_k$  as in (152),  $A$  as in (153). Here*

$$A_0(R_2) := \left( \int_{R_1}^{R_2} (A(w))^{p/(\lambda_\alpha + \lambda_\beta)} dw \right)^{\frac{\lambda_\alpha + \lambda_\beta}{p}}, \quad (161)$$

$$\tilde{\gamma}_1 := \begin{cases} 2^{\left(\frac{\lambda_\alpha + \lambda_\beta}{\lambda_\nu}\right)} - 1, & \text{if } \lambda_\alpha + \lambda_\beta \geq \lambda_\nu, \\ 1, & \text{if } \lambda_\alpha + \lambda_\beta \leq \lambda_\nu. \end{cases} \quad (162)$$

Put

$$\begin{aligned} \varphi_3(R_2) := A_0(R_2) & \left( \frac{\lambda_\nu}{(\lambda_\alpha + \lambda_\beta)p} \right)^{(\lambda_\nu/p)} \\ & \left[ \lambda_\alpha^{(\lambda_\nu/p)} + 2 \left( \frac{\lambda_\alpha + \lambda_\beta}{p} \right) (\tilde{\gamma}_1 \lambda_\beta)^{(\frac{\lambda_\nu}{p})} \right]. \end{aligned} \quad (163)$$

Then

$$\begin{aligned} & \int_A \left[ \sum_{j=1}^{M-1} \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_j(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_2} f_{j+1}(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \right. \\ & \quad \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_j(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^{\gamma_1} f_{j+1}(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_{j+1}(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] + \\ & \quad \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_2} f_M(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \\ & \quad \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_1(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^{\gamma_1} f_M(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_M(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] dx \leq \\ & 2\varphi_3(R_2) \left[ \sum_{j=1}^M \left( \int_A \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^p dx \right) \right]. \end{aligned} \quad (164)$$

**Proof.** As in Theorem 68, based on Theorem 47.  $\square$

We show the special important case next.

**Theorem 78.** Let all as in Assumption 74 without  $\lambda_\nu$  there. Here  $\gamma_2 = \gamma_1 + 1$ , and let  $\lambda_\alpha > 0$ ,  $\lambda_\beta := \lambda_{\alpha+1}$ ,  $0 < \lambda_{\alpha+1} < 1$ , such that  $p := \lambda_\alpha + \lambda_{\alpha+1} > 1$ . Denote

$$\theta_3 := \begin{cases} 2^{(\lambda_\alpha/\lambda_{\alpha+1})} - 1, & \text{if } \lambda_\alpha \geq \lambda_{\alpha+1}, \\ 1, & \text{if } \lambda_\alpha \leq \lambda_{\alpha+1}, \end{cases} \quad (165)$$

$$\begin{aligned} L(R_2) &:= \left[ 2 \frac{(1 - \lambda_{\alpha+1})}{(N - \lambda_{\alpha+1})} \right. \\ & \quad \left. \left( R_2^{\frac{N - \lambda_{\alpha+1}}{1 - \lambda_{\alpha+1}}} - R_1^{\frac{N - \lambda_{\alpha+1}}{1 - \lambda_{\alpha+1}}} \right) \right]^{(1 - \lambda_{\alpha+1})} \left( \frac{\theta_3 \lambda_{\alpha+1}}{\lambda_\alpha + \lambda_{\alpha+1}} \right)^{\lambda_{\alpha+1}}, \end{aligned} \quad (166)$$

and

$$P(R_2) := \int_{R_1}^{R_2} (R_2 - t)^{(\nu - \gamma_1 - 1)(\frac{p}{p-1})} t^{(\frac{1-N}{p-1})} dt, \quad (167)$$

$$\Phi(R_2) := L(R_2) \left( \frac{P_1(R_2)^{(p-1)}}{(\Gamma(\nu - \gamma_1))^p} \right)^{2^{(p-1)}}. \quad (168)$$

Then

$$\begin{aligned}
& \int_A \left\{ \left\{ \sum_{j=1}^{M-1} \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_j(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_1+1} f_{j+1}(x)}{\partial r^{\gamma_1+1}} \right|^{\lambda_{\alpha+1}} + \right. \right. \right. \\
& \quad \left. \left. \left| \frac{\partial_{*R_1}^{\gamma_1} f_{j+1}(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_1+1} f_j(x)}{\partial r^{\gamma_1+1}} \right|^{\lambda_{\alpha+1}} \right] \right\} + \\
& \quad \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_1+1} f_M(x)}{\partial r^{\gamma_1+1}} \right|^{\lambda_{\alpha+1}} + \right. \\
& \quad \left. \left| \frac{\partial_{*R_1}^{\gamma_1} f_M(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_1+1} f_1(x)}{\partial r^{\gamma_1+1}} \right|^{\lambda_{\alpha+1}} \right] \Big\} dx \leq \\
& 2\Phi(R_2) \left[ \sum_{j=1}^M \left( \int_A \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^p dx \right) \right]. \tag{169}
\end{aligned}$$

**Proof.** As in Theorem 68, based on Theorem 48.  $\square$

We study the  $L_\infty$  case next.

We need to make

**Assumption 79.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, 2$ ,  $n := \lceil \nu \rceil$ ,  $f_j : \bar{A} \rightarrow \mathbb{R}$  with  $f_j \in L_1(A)$ ,  $j = 1, \dots, M$ ,  $M \in \mathbb{N}$ , where  $\mathbb{R}^N \supseteq A := B(0, R_2) - \bar{B}(0, R_1)$ ,  $0 < R_1 < R_2$ ,  $N \geq 2$ . Assume that  $f_j(\cdot\omega) \in AC^n([R_1, R_2])$  for all  $\omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f_j(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ , for all  $\omega \in S^{N-1}$ ;  $j = 1, \dots, M$ . Further assume that  $\frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \in L_\infty(A)$ ,  $j = 1, \dots, M$ . More precisely, for these  $r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ , for which  $D_{*R_1}^\nu f_j(r\omega)$  takes real values, there exists  $M_j > 0$  such that

$$|D_{*R_1}^\nu f_j(r\omega)| \leq M_j, \text{ for } j = 1, \dots, M. \tag{170}$$

We suppose that

$$\frac{\partial^k f_j(R_1\omega)}{\partial r^k} = 0, k = 0, 1, \dots, n-1,$$

$\forall \omega \in S^{N-1}$ ;  $j = 1, \dots, M$ .

Let  $\lambda_\nu, \lambda_\alpha, \lambda_\beta \geq 0$ . If  $\gamma_1 = 0$  we set  $\lambda_\alpha = 1$  and if  $\gamma_2 = 0$  we set  $\lambda_\beta = 1$ .

The last main result follows.

**Theorem 80.** All as in Assumption 79.

Set

$$\begin{aligned}
T(R_2) &:= \frac{R_2^{N-1}}{(\nu\lambda_\alpha - \gamma_1\lambda_\alpha + \nu\lambda_\beta - \gamma_2\lambda_\beta + 1)} \cdot \\
& \frac{(R_2 - R_1)^{(\nu\lambda_\alpha - \gamma_1\lambda_\alpha + \nu\lambda_\beta - \gamma_2\lambda_\beta + 1)}}{(\Gamma(\nu - \gamma_1 + 1))^{\lambda_\alpha} (\Gamma(\nu - \gamma_2 + 1))^{\lambda_\beta}}. \tag{171}
\end{aligned}$$

Then

$$\begin{aligned}
& \int_A \left\{ \left[ \sum_{j=1}^{M-1} \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_j(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_2} f_{j+1}(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_j(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \right. \right. \\
& \quad \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_j(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^{\gamma_1} f_{j+1}(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_{j+1}(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] \right\} + \\
& \quad \left[ \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^{\gamma_2} f_M(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} + \right. \\
& \quad \left. \left| \frac{\partial_{*R_1}^{\gamma_2} f_1(x)}{\partial r^{\gamma_2}} \right|^{\lambda_\beta} \left| \frac{\partial_{*R_1}^{\gamma_1} f_M(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_M(x)}{\partial r^\nu} \right|^{\lambda_\nu} \right] \Bigg\} dx \leq \\
& \quad \frac{2\pi^{N/2}}{\Gamma(N/2)} T(R_2) \left\{ \sum_{j=1}^{M-1} \left\{ M_j^{2(\lambda_\alpha + \lambda_\nu)} + M_j^{2\lambda_\beta} \right\} \right\}. \tag{172}
\end{aligned}$$

**Proof.** Based on Theorem 52; here  $p(r) = r^{N-1}$ ,  $r \in [R_1, R_2]$ , apply (97)  $\forall \omega \in S^{N-1}$ . It goes as the proof of Theorem 66. Finally use Theorem 64 and Proposition 60.  $\square$

## 6 Applications

We need

**Corollary 81.** (to Theorem 68,  $f_1 = f_2$ ) All as in Theorem 68. It holds

$$\begin{aligned}
& \int_A \left| \frac{\partial_{*R_1}^{\gamma_1} f_1(x)}{\partial r^{\gamma_1}} \right|^{\lambda_\alpha} \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^{\lambda_\nu} dx \leq \\
& (A_0(R_2) |_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{p} \right)^{\left( \frac{\lambda_\nu}{p} \right)} \left( \int_A \left| \frac{\partial_{*R_1}^\nu f_1(x)}{\partial r^\nu} \right|^p dx \right). \tag{173}
\end{aligned}$$

So setting  $\lambda_\alpha = \lambda_\nu = 1$ ,  $p = 2$  in (173), we obtain in detail

**Proposition 82.** Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f : \bar{A} \rightarrow \mathbb{R}$  with  $f \in L_1(A)$ , where  $A := B(0, R_2) - \bar{B}(0, R_1) \subseteq \mathbb{R}^N$ ,  $N \geq 2$ ,  $0 < R_1 < R_2$ . Assume that  $f(\cdot\omega) \in AC^n([R_1, R_2])$ ,  $\forall \omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ ,  $\forall \omega \in S^{N-1}$ . Further assume that  $\frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \in L_\infty(A)$ . More precisely, for these  $r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ , for which  $D_{*R_1}^\nu f(r\omega)$  takes real values,  $\exists M_1 > 0$  such that

$$|D_{*R_1}^\nu f(r\omega)| \leq M_1.$$

Suppose that

$$\frac{\partial^j f(R_1 \omega)}{\partial r^j} = 0, j = 0, 1, \dots, n-1, \forall \omega \in S^{N-1}.$$

Set

$$P(r) := \int_{R_1}^r (r-t)^{2(\nu-\gamma-1)} t^{(1-N)} dt, \quad R_1 \leq r \leq R_2, \quad (174)$$

$$A(r) := \frac{r^{\left(\frac{N-1}{2}\right)} \sqrt{P(r)}}{\Gamma(\nu-\gamma)}, \quad (175)$$

$$\tilde{A}_0(R_2) := \left( \int_{R_1}^{R_2} (A(r))^2 dr \right)^{1/2}. \quad (176)$$

Then

$$\begin{aligned} & \int_A \left| \frac{\partial_{*R_1}^\gamma f(x)}{\partial r^\gamma} \right| \left| \frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \right| dx \leq \\ & \tilde{A}_0(R_2) 2^{-1/2} \left( \int_A \left( \frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \right)^2 dx \right). \end{aligned} \quad (177)$$

When  $\gamma = 0$  we get in detail

**Proposition 83.** Let  $\nu \geq 1$ ,  $n := \lceil \nu \rceil$ ,  $f : \bar{A} \rightarrow \mathbb{R}$  with  $f \in L_1(A)$ , where  $A := B(0, R_2) - \bar{B}(0, R_1) \subseteq \mathbb{R}^N$ ,  $N \geq 2$ ,  $0 < R_1 < R_2$ . Assume that  $f(\cdot \omega) \in AC^n([R_1, R_2])$ ,  $\forall \omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f(\cdot \omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ ,  $\forall \omega \in S^{N-1}$ . Further assume that  $\frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \in L_\infty(A)$ . More precisely, for these  $r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ , for which  $D_{*R_1}^\nu f(r\omega)$  takes real values,  $\exists M_1 > 0$  such that

$$|D_{*R_1}^\nu f(r\omega)| \leq M_1.$$

Suppose that

$$\frac{\partial^j f(R_1 \omega)}{\partial r^j} = 0, j = 0, 1, \dots, n-1, \forall \omega \in S^{N-1}.$$

Set

$$P_0(r) := \int_{R_1}^r (r-t)^{2(\nu-1)} t^{(1-N)} dt, \quad R_1 \leq r \leq R_2, \quad (178)$$

$$A_*(r) := \frac{r^{\left(\frac{N-1}{2}\right)} \sqrt{P_0(r)}}{\Gamma(\nu)}, \quad (179)$$

$$\tilde{A}_0(R_2) := \left( \int_{R_1}^{R_2} (A_*(r))^2 dr \right)^{1/2}. \quad (180)$$

Then

$$\int_A |f(x)| \left| \frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \right| dx \leq$$

$$\tilde{A}_0(R_2) 2^{-1/2} \left( \int_A \left( \frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \right)^2 dx \right). \quad (181)$$

Based on Corollary 35 we give

**Proposition 84.** Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f : \bar{A} \rightarrow \mathbb{R}$  with  $f \in L_1(A)$ , where  $A := B(0, R_2) - \bar{B}(0, R_1) \subseteq \mathbb{R}^N$ ,  $N \geq 2$ ,  $0 < R_1 < R_2$ . Assume that  $f(\cdot\omega) \in AC^n([R_1, R_2])$ ,  $\forall \omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ ,  $\forall \omega \in S^{N-1}$ .

Suppose that

$$\frac{\partial^j f(R_1\omega)}{\partial r^j} = 0, \quad j = 0, 1, \dots, n-1, \quad \forall \omega \in S^{N-1}.$$

Then

$$\begin{aligned} 1) \quad & \int_{R_1}^r |D_{*R_1}^\gamma f(t\omega)| |D_{*R_1}^\nu f(t\omega)| dt \leq \\ & \left( \frac{(r - R_1)^{(\nu-\gamma)}}{2\Gamma(\nu-\gamma) \sqrt{\nu-\gamma} \sqrt{2\nu-2\gamma-1}} \right) \\ & \left( \int_{R_1}^r (D_{*R_1}^\nu f(t\omega))^2 dt \right), \quad \text{all } R_1 \leq r \leq R_2, \quad \forall \omega \in S^{N-1}. \end{aligned} \quad (182)$$

2) When  $\gamma = 0$  we get

$$\begin{aligned} & \int_{R_1}^r |f(t\omega)| |D_{*R_1}^\nu f(t\omega)| dt \leq \\ & \left( \frac{(r - R_1)^\nu}{2\Gamma(\nu) \sqrt{\nu} \sqrt{2\nu-1}} \right) \left( \int_{R_1}^r (D_{*R_1}^\nu f(t\omega))^2 dt \right), \end{aligned} \quad (183)$$

all  $R_1 \leq r \leq R_2$ ,  $\forall \omega \in S^{N-1}$ .

In particular we have

$$\begin{aligned} 3) \quad & \int_{R_1}^{R_2} |f(r\omega)| |D_{*R_1}^\nu f(r\omega)| dr \leq \\ & \left( \frac{(R_2 - R_1)^\nu}{2\Gamma(\nu) \sqrt{\nu} \sqrt{2\nu-1}} \right) \left( \int_{R_1}^{R_2} (D_{*R_1}^\nu f(r\omega))^2 dr \right), \quad \forall \omega \in S^{N-1}. \end{aligned} \quad (184)$$

Next we apply Proposition 84, see (183) for proving uniqueness of solution in a PDE initial value problem on A.

**Theorem 85.** Let  $\nu > 1$ ,  $\nu \notin \mathbb{N}$ ,  $n := \lceil \nu \rceil$ ,  $f : \bar{A} \rightarrow \mathbb{R}$  with  $f \in L_1(A)$ , where  $A := B(0, R_2) - \bar{B}(0, R_1) \subseteq \mathbb{R}^N$ ,  $N \geq 2$ ,  $0 < R_1 < R_2$ . Assume that  $f(\cdot\omega) \in AC^n([R_1, R_2])$ ,  $\forall \omega \in S^{N-1}$ , and that  $\frac{\partial_{*R_1}^\nu f(\cdot\omega)}{\partial r^\nu} \in AC([R_1, R_2])$ ,  $\forall \omega \in S^{N-1}$ . Further assume  $\frac{D_{*R_1}^\nu f(x)}{\partial r^\nu} \in L_\infty(A)$ , such that there exists  $M_1 > 0$  with  $|D_{*R_1}^\nu f(r\omega)| \leq M_1$ ,  $\forall r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ . Suppose that

$$\frac{\partial^j f(R_1\omega)}{\partial r^j} = 0, \quad j = 0, 1, \dots, n-1, \quad \forall \omega \in S^{N-1}.$$



Consider the PDE

$$\frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu f(x)}{\partial r^\nu} \right) = \theta(x) f(x), \quad (185)$$

$\forall x \in \bar{A}$ , where  $0 \neq \theta : \bar{A} \rightarrow \mathbb{R}$  is continuous. If (185) has a solution then it is unique.

**Proof.** We rewrite (185) as

$$\frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu f(r\omega)}{\partial r^\nu} \right) = \theta(r\omega) f(r\omega), \quad (186)$$

valid  $\forall r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$  and  $0 \neq \theta : ([R_1, R_2] \times S^{N-1}) \rightarrow \mathbb{R}$  is continuous.

Assume  $f_1$  and  $f_2$  are solution to (185), then

$$\frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu f_1(r\omega)}{\partial r^\nu} \right) = \theta(r\omega) f_1(r\omega), \quad (187)$$

and

$$\frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu f_2(r\omega)}{\partial r^\nu} \right) = \theta(r\omega) f_2(r\omega), \quad (188)$$

$\forall r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ .

Call  $g := f_1 - f_2$ , thus by subtraction in (187) we get

$$\frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu g(r\omega)}{\partial r^\nu} \right) = \theta(r\omega) g(r\omega), \quad (189)$$

$\forall r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ . Of course

$$\frac{\partial^j g(R_1\omega)}{\partial r^j} = 0, \quad j = 0, 1, \dots, n-1, \quad \forall \omega \in S^{N-1}.$$

Consequently we have

$$\begin{aligned} & \left( \frac{\partial_{*R_1}^\nu g(r\omega)}{\partial r^\nu} \right) \frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu g(r\omega)}{\partial r^\nu} \right) \\ &= \theta(r\omega) g(r\omega) \left( \frac{\partial_{*R_1}^\nu g(r\omega)}{\partial r^\nu} \right), \end{aligned} \quad (190)$$

$\forall r \in [R_1, R_2]$ ,  $\forall \omega \in S^{N-1}$ .

Hence

$$\begin{aligned} & \int_{R_1}^r \left( \frac{\partial_{*R_1}^\nu g(t\omega)}{\partial r^\nu} \right) \frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu g(t\omega)}{\partial r^\nu} \right) dt \\ &= \int_{R_1}^r \theta(t\omega) g(t\omega) \left( \frac{\partial_{*R_1}^\nu g(t\omega)}{\partial r^\nu} \right) dt, \end{aligned} \quad (191)$$

$\forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ .

Therefore we find

$$\left. \frac{\left( \frac{\partial_{*R_1}^\nu g(t\omega)}{\partial r^\nu} \right)^2}{2} \right|_{R_1}^r = \int_{R_1}^r \theta(t\omega) g(t\omega) \left( \frac{\partial_{*R_1}^\nu g(t\omega)}{\partial r^\nu} \right) dt, \quad (192)$$

$\forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ .

Notice that  $\frac{\partial_{*R_1}^\nu g(R_1\omega)}{\partial r^\nu} = 0, \forall \omega \in S^{N-1}$ , see (110).

Consequently we find

$$\begin{aligned} \left( \frac{\partial_{*R_1}^\nu g(r\omega)}{\partial r^\nu} \right)^2 &= 2 \left| \int_{R_1}^r \theta(t\omega) g(t\omega) \left( \frac{\partial_{*R_1}^\nu g(t\omega)}{\partial r^\nu} \right) dt \right| \\ &\leq 2 \|g\|_\infty \int_{R_1}^r |g(t\omega)| \left| \frac{\partial_{*R_1}^\nu g(t\omega)}{\partial r^\nu} \right| dt \end{aligned} \quad (193)$$

$$\stackrel{(183)}{\leq} \left( \frac{\|\theta\|_\infty (r - R_1)^\nu}{\Gamma(\nu) \sqrt{\nu} \sqrt{2\nu - 1}} \right) \quad (194)$$

$$\begin{aligned} &\left( \int_{R_1}^r (D_{*R_1}^\nu g(t\omega))^2 dt \right) \leq \\ &\left( \frac{\|\theta\|_\infty (R_2 - R_1)^\nu}{\Gamma(\nu) \sqrt{\nu} \sqrt{2\nu - 1}} \right) \left( \int_{R_1}^r (D_{*R_1}^\nu g(t\omega))^2 dt \right), \end{aligned} \quad (195)$$

$\forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ .

Call

$$K := \frac{\|\theta\|_\infty (R_2 - R_1)^\nu}{\Gamma(\nu) \sqrt{\nu} \sqrt{2\nu - 1}} > 0. \quad (196)$$

So we have proved that

$$\left( \frac{\partial_{*R_1}^\nu g(r\omega)}{\partial r^\nu} \right)^2 \leq K \left( \int_{R_1}^r (D_{*R_1}^\nu g(t\omega))^2 dt \right), \quad (197)$$

$\forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ . Here  $D_{*R_1}^\nu g(\cdot\omega) \in C([R_1, R_2]), \forall \omega \in S^{N-1}$ .

Hence by Grönwall's inequality we get  $(D_{*R_1}^\nu g(r\omega))^2 \equiv 0$ , so that  $D_{*R_1}^\nu g(r\omega) \equiv 0, \forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ . Thus  $\frac{\partial}{\partial r} \left( \frac{\partial_{*R_1}^\nu g(r\omega)}{\partial r^\nu} \right) \equiv 0, \forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ . And by (188) we have  $\theta(r\omega) g(r\omega) \equiv 0, \forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ , implying  $g(r\omega) \equiv 0, \forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ .

Hence proving  $f_1(r\omega) = f_2(r\omega), \forall r \in [R_1, R_2], \forall \omega \in S^{N-1}$ . Thus

$$f_1(x) = f_2(x), \forall x \in \bar{A},$$

hence proving the claim.  $\square$

We give the very important

**Remark 86.** From Corollary 12 we saw that: for  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f \in AC^n([a, b])$ , given that  $D_a^\nu f(x)$  exists in  $\mathbb{R}$ ,  $\forall x \in [a, b]$ , and  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ , imply that  $D_{*a}^\nu f = D_a^\nu f$ . Also we saw in Theorem 16, 17, that by adding to the assumptions of Theorem 16 that "there exists  $D_a^\nu f(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$ ", we can rewrite the conclusions of Theorem 16, that is getting the conclusions of Theorem 17, in the language of Riemann-Liouville fractional derivatives. Notice there, under the above additional assumption that also holds  $D_a^\gamma f(x) = D_{*a}^\gamma f(x)$ ,  $\forall x \in [a, b]$ .

Theorem 16 is where is based the whole article.

So by adding to the assumptions of all of our results here for all functions involved that "there exists  $D_a^\nu f(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$ " we can rewrite them all in terms of Riemann-Liouville fractional derivatives. Accordingly for the case of spherical shell we need to add "there exists  $D_{R_1}^\nu f(r\omega) \in \mathbb{R}$ ,  $\forall r \in [R_1, R_2]$ , for each  $\omega \in S^{N-1}$ ", and all can be rewritten in terms of Riemann-Liouville radial fractional derivatives.

So as examples next we present only few of all these can be rewritten results.

We present

**Theorem 87.** Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ . Call  $n := \lceil \nu \rceil$  and assume  $f \in AC^n([a, b])$  such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ , and  $\exists D_a^\nu f(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$  with  $D_a^\nu f \in L_\infty(a, b)$ . Let  $p, q > 1$  such that  $\frac{1}{p} + \frac{1}{q} = 1$ ,  $a \leq x \leq b$ .

Then

$$\begin{aligned} & \int_a^x |D_a^\gamma f(\omega)| |D_a^\nu f(\omega)| d\omega \leq \\ & \frac{(x-a)^{\frac{p\nu-p\gamma-p+2}{p}}}{(\sqrt[p]{2}) \Gamma(\nu-\gamma) ((p\nu-p\gamma-p+1)(p\nu-p\gamma-p+2))^{1/p}} \\ & \cdot \left( \int_a^x |D_a^\nu f(\omega)|^q d\omega \right)^{2/q}. \end{aligned} \quad (198)$$

**Proof.** Similar to Theorem 18.  $\square$

The converse result follows.

**Theorem 88.** Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ . Call  $n := \lceil \nu \rceil$  and assume  $f \in AC^n([a, b])$  such that  $f^{(k)}(a) = 0$ ,  $k = 0, 1, \dots, n-1$ , and  $\exists D_a^\nu f(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$  with  $D_a^\nu f, \frac{1}{D_a^\nu f} \in L_\infty(a, b)$ . Suppose that  $D_a^\nu f$  is of fixed sign a.e in  $[a, b]$ . Let  $p, q$  such that  $0 < p < 1$ ,  $q < 0$  and  $\frac{1}{p} + \frac{1}{q} = 1$ ,  $a \leq x \leq b$ .

Then

$$\begin{aligned} & \int_a^x |D_a^\gamma f(\omega)| |D_a^\nu f(\omega)| d\omega \geq \\ & \frac{(x-a)^{\frac{p\nu-p\gamma-p+2}{p}}}{(\sqrt[p]{2}) \Gamma(\nu-\gamma) ((p\nu-p\gamma-p+1)(p\nu-p\gamma-p+2))^{1/p}} \\ & \cdot \left( \int_a^x |D_a^\nu f(\omega)|^q d\omega \right)^{2/q}. \end{aligned} \quad (199)$$

**Proof.** As in Theorem 20.  $\square$

We present

**Theorem 89.** Let  $\nu \geq \gamma_i + 1$ ,  $\gamma_i \geq 0$ ,  $i = 1, 2$ ,  $n := \lceil \nu \rceil$ , and  $f_1, f_2 \in AC^n([a, b])$  such that  $f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ . Consider also  $p(t) > 0$  and  $q(t) \geq 0$ , with all  $p(t)$ ,  $\frac{1}{p(t)}$ ,  $q(t) \in L_\infty(a, b)$ . Further assume  $\exists D_a^\nu f_i(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$  and  $D_a^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ . Let  $\lambda_\nu > 0$  and  $\lambda_\alpha, \lambda_\beta \geq 0$  such that  $\lambda_\nu < p$ , where  $p > 1$ .

Here  $P_k$  is as in (39),  $A(\omega)$  is as in (40),  $A_0(x)$  as in (41),  $\delta_1$  as in (42). If  $\lambda_\beta = 0$ , we obtain that

$$\begin{aligned} & \int_a^x q(\omega) \left[ |D_a^{\gamma_1} f_1(\omega)|^{\lambda_\alpha} |D_a^\nu f_1(\omega)|^{\lambda_\nu} + \right. \\ & \quad \left. |D_a^{\gamma_1} f_2(\omega)|^{\lambda_\alpha} |D_a^\nu f_2(\omega)|^{\lambda_\nu} \right] d\omega \leq \\ & \quad (A_0(x)|_{\lambda_\beta=0}) \left( \frac{\lambda_\nu}{\lambda_\alpha + \lambda_\nu} \right)^{\lambda_\nu/p} \delta_1 \\ & \quad \left[ \int_a^x p(\omega) [|D_a^\nu f_1(\omega)|^p + |D_a^\nu f_2(\omega)|^p] d\omega \right]^{\left( \frac{\lambda_\alpha + \lambda_\nu}{p} \right)}. \end{aligned} \quad (200)$$

**Proof.** As in Theorem 30.  $\square$

**Corollary 90.** (All as in Theorem 89,  $\lambda_\beta = 0$ ,  $p(t) = q(t) = 1$ ,  $\lambda_\alpha = \lambda_\nu = 1$ ,  $p = 2$ .) In detail:

Let  $\nu \geq \gamma_1 + 1$ ,  $\gamma_1 \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f_1, f_2 \in AC^n([a, b]) : f_1^{(j)}(a) = f_2^{(j)}(a) = 0$ ,  $j = 0, 1, \dots, n-1$ ,  $a \leq x \leq b$ ;  $\exists D_a^\nu f_i(x) \in \mathbb{R}$ ,  $\forall x \in [a, b]$  with  $D_a^\nu f_i \in L_\infty(a, b)$ ,  $i = 1, 2$ . Then

$$\begin{aligned} & \int_a^x [|D_a^{\gamma_1} f_1(\omega)| |(D_a^\nu f_1)(\omega)| + \\ & \quad |(D_a^{\gamma_1} f_2)(\omega)| |(D_a^\nu f_2)(\omega)|] d\omega \leq \\ & \quad \left( \frac{(x-a)^{(\nu-\gamma_1)}}{2\Gamma(\nu-\gamma_1)\sqrt{\nu-\gamma_1}\sqrt{2\nu-2\gamma_1-1}} \right) \\ & \quad \left( \int_a^x [(D_a^\nu f_1(\omega))^2 + (D_a^\nu f_2(\omega))^2] d\omega \right), \end{aligned} \quad (201)$$

all  $a \leq x \leq b$ .

We need

**Definition 91.** Let  $F : \bar{A} \rightarrow \mathbb{R}$ ,  $\nu \geq 0$ ,  $n := \lceil \nu \rceil$  such that  $F(\cdot\omega) \in AC^n([R_1, R_2])$ , for all  $\omega \in S^{N-1}$ . We call the Riemann-Liouville radial fractional derivative the following function

$$\frac{\partial_{R_1}^\nu F(x)}{\partial r^\nu} := \frac{1}{\Gamma(n-\nu)} \frac{\partial^n}{\partial r^n} \int_{R_1}^r (r-t)^{n-\nu-1} F(t\omega) dt, \quad (202)$$

where  $x \in \bar{A}$ , i.e.  $x = r\omega$ ,  $r \in [R_1, R_2]$ ,  $\omega \in S^{N-1}$ .

Clearly

$$\frac{\partial_{*R_1}^0 F(x)}{\partial r^0} = F(x),$$

and

$$\frac{\partial_{R_1}^\nu F(x)}{\partial r^\nu} = \frac{\partial^\nu F(x)}{\partial r^\nu}, \text{ if } \nu \in \mathbb{N}.$$

We give

**Proposition 92.** Let  $\nu \geq \gamma + 1$ ,  $\gamma \geq 0$ ,  $n := \lceil \nu \rceil$ ,  $f : \bar{A} \rightarrow \mathbb{R}$  with  $f \in L_1(A)$ , where  $A := B(0, R_2) - B(0, R_1) \subseteq \mathbb{R}^N$ ,  $N \geq 2$ ,  $0 < R_1 < R_2$ . Assume that  $f(\cdot\omega) \in AC^n([R_1, R_2])$ ,  $\forall \omega \in S^{N-1}$ , and that  $\frac{\partial_{R_1}^\nu f(\cdot\omega)}{\partial r^\nu} \in L_\infty(R_1, R_2)$ ,  $\forall \omega \in S^{N-1}$ . Further assume that  $\exists D_{R_1}^\nu f(r\omega) \in \mathbb{R}$ ,  $\forall r \in [R_1, R_2]$ , for each  $\omega \in S^{N-1}$ , with  $\frac{\partial_{R_1}^\nu f(x)}{\partial r^\nu} \in L_\infty(A)$ . We suppose  $\forall r \in [R_1, R_2]$  and  $\forall \omega \in S^{N-1}$  that  $\exists M_1 > 0$  such that  $|D_{R_1}^\nu f(r\omega)| \leq M_1$ .

Suppose that

$$\frac{\partial^j f(R_1\omega)}{\partial r^j} = 0, j = 0, 1, \dots, n-1, \forall \omega \in S^{N-1}.$$

Set

$$P(r) := \int_{R_1}^r (r-t)^{2(\nu-\gamma-1)} t^{(1-N)} dt, \quad R_1 \leq r \leq R_2, \quad (203)$$

also

$$A(r) := \frac{r^{\left(\frac{N-1}{2}\right)} \sqrt{P(r)}}{\Gamma(\nu-\gamma)}, \quad (204)$$

$$\tilde{A}_0(R_2) := \left( \int_{R_1}^{R_2} (A(r))^2 dr \right)^{1/2}. \quad (205)$$

Then

$$\begin{aligned} \int_A \left| \frac{\partial_{R_1}^\gamma f(x)}{\partial r^\gamma} \right| \left| \frac{\partial_{R_1}^\nu f(x)}{\partial r^\nu} \right| dx &\leq \\ \tilde{A}_0(R_2) 2^{-1/2} \left( \int_A \left( \frac{\partial_{R_1}^\nu f(x)}{\partial r^\nu} \right)^2 dx \right). \end{aligned} \quad (206)$$

**Proof.** Similarly as in Proposition 82.  $\square$

## References

- [1] Aliprantis, Charalambos D., Burkinshaw, Owen, *Principles of Real Analysis*, Third Edition, Academic Press, Inc., San Diego, CA, 1998.
- [2] G. A. Anastassiou, *Quantitative Approximations*, Chapman & Hall/CRC, Boca Raton, New York, 2001.
- [3] G. A. Anastassiou, *Opial type inequalities involving fractional derivatives of two functions and applications*, Computers and Mathematics with Applications, 48 (2004), 1701-1731.
- [4] Anastassiou, George A. *Fractional Opial inequalities for several functions with applications*, J. Comput. Anal. Appl. 7 (2005), no. 3, 233-259.
- [5] G. A. Anastassiou, *Riemann-Liouville fractional multivariate Opial type inequalities on spherical shells*, accepted for publication, Bulletin of Allahabad Math. Soc., India, 2007.
- [6] G. A. Anastassiou, *Fractional Multivariate Opial type inequalities over spherical shells*, Communications in Applied Analysis 11 (2007), no. 2, 201-233.
- [7] G. A. Anastassiou, *Reverse Riemann-Liouville fractional Opial inequalities for several functions*, Submitted 2007.
- [8] G. A. Anastassiou, *Converse fractional Opial inequalities for several functions*, Submitted 2007.
- [9] G. A. Anastassiou, *Opial Type Inequalities Involving Riemann-Liouville Fractional derivatives of two functions*, accepted for publication, Mathematics and Computer Modelling, 2007.
- [10] G. A. Anastassiou, *Riemann-Liouville Fractional Opial inequalities for several functions with Applications*, accepted , International I. of Pure and Appl. Math.,2007.
- [11] G. A. Anastassiou, J. J. Koliha and J. Pecaric, *Opial inequalities for fractional derivatives*, Dynam. Systems Appl. 10 (2001), no. 3, 395-406.
- [12] G. A. Anastassiou, J. J. Koliha and J. Pecaric, *Opial type  $L^p$  inequalities for fractional derivatives*, Intern. Journal of Mathematics and Math. Sci., Vol. 31, no. 2 (2002), 85-95.
- [13] P. R. Bessack, *On an integral inequality of Z. Opial*, Trans. Amer. Math. Soc., 104 (1962), 470-475.
- [14] Caputo M. (1967), *Linear Models of Dissipation Whose  $Q$  is Almost Frequency Independent-II*, Geophys J Royal Astronom Soc 13:529-539.

- [15] Caputo M., Mainardi F. (1971a), *A New Dissipation Model Based on Memory Mechanism*, Pure and Appl Geophys 91:134-147.
- [16] Caputo M., Mainardi F. (1971b), *Linear Models of Dissipation in Anelastic Solid*, Rivista del Nuovo Cimento 1:161-198.
- [17] Kai Diethelm, *Fractional Differential Equations*, on line: <http://www.tu-bs.de/~diethelm/lehre/f-dgl02/fde-skript.ps.gz>
- [18] G. D. Handley, J. J. Koliha and J. Pecaric, *Hilbert-Pachpatte type integral inequalities for fractional derivatives*, Fract. Calc. Appl. Anal. 4, Vol 1 (2001), 37-46.
- [19] Virginia Kiryakova, *Generalized fractional calculus and applications*, Pitman Research Notes in Math. Series, 301. Longman Scientific and Technical, Harlow; co-published in U.S.A with John Wiley and Sons, Inc., New York, 1994.
- [20] Kenneth Miller, B. Ross, *An Introduction to the Fractional Calculus and fractional differential equations*, John Wiley and Sons, Inc., New York, 1993.
- [21] Keith Oldham, Jerome Spanier, *The Fractional Calculus: Theory and Applications of Differentiation and Integration to arbitrary order*, Dover Publications, New York, 2006.
- [22] Z. Opial, *Sur une inégalité*, Ann. Polon. Math. 8 (1960), 29-32.
- [23] H. L. Royden, *Real Analysis*, second edition, Macmillan, 1968, New York.
- [24] W. Rudin, *Real and Complex Analysis*, International Student Edition, McGraw Hill 1970, London, New York.
- [25] D. Stroock, *A Concise Introduction to the Theory of Integration*, 3rd Edition, Birkhäuser, Boston, Basel, Berlin, 1999.
- [26] E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis*, Cambridge University Press, 1927.
- [27] D. Willett, *The existence-uniqueness theorem for an  $n$ th order linear ordinary differential equation*, Amer. Math. Monthly, 75 (1968), 174-178.

## Asymptotics for Szegő polynomials with respect to a class of weakly convergent measures

Michael Arciero<sup>1</sup>, Lewis Pakula<sup>2</sup>

<sup>1</sup>*University of New England, Biddeford ME, 04005, marciero@une.edu*

<sup>2</sup>*University of Rhode Island, Kingston RI, 02881, pakula@math.uri.edu*

### Abstract

Recent results of the author characterize limits for *Szegő* polynomials of fixed degree  $k$  with respect to measures which are weakly convergent to a sum of  $m < k$  point masses, with the measures formed by convolving the point masses with the Poisson and Fejér kernels. Moreover, the limit polynomial is seen to be the same in each case. Here, we show that the Poisson kernel can be expressed as a convex combination of Fejér kernels. Conjectures are made for a general class of kernels whose Fourier coefficients  $\widehat{\mu}(\ell)$  form convex functions of  $\ell$ .

*Keywords:* Szegő polynomial, orthogonal polynomial, frequency analysis, Poisson kernel, Fejér kernel.

## 1 Introduction

Given a measure,  $\mu$ , on the unit circle, the Szegő polynomial of degree  $k$  with respect to  $\mu$ , which we denote  $P_k(z, \mu)$ , is the polynomial in the complex variable  $z$  which attains the minimum

$$\min_{p \in \Lambda_k} \int_{-\pi}^{\pi} |p(e^{i\theta})|^2 d\mu(\theta) = \int_{-\pi}^{\pi} |P_k(e^{i\theta}, \mu)|^2 d\mu(\theta), \quad (1)$$

where  $\Lambda_k$  is the set of monic polynomials of degree  $k$ . The Szegő polynomials with respect to a measure  $\mu$  form an orthogonal sequence, are uniquely defined if the degree is less than the number of points on which  $\mu$  is supported, and can be expressed as a ratio of matrix determinants or generated recursively using *Levinson's recursion*. Szegő polynomials have many applications and have been studied widely. See [4, 6, 10, 11] for background. Some results related to frequency analysis appear in [5, 7, 9, 8]. The motivation for the use of Szegő polynomials in frequency analysis is loosely based on the observation that the spectral measure of a digital signal with strong sinusoidal components will be heavily weighted at the frequency locations  $\theta_j$ , and in light of (1), one would expect  $P_k(z, \mu)$  to have zeros near  $e^{i\theta_j}$ . Note that for



Table 1: Poisson and Fejér kernels

kernel	density	moments $\widehat{\psi}_h(\ell)$	$h$
Poisson:	$\psi_h(\theta) = \frac{1-r^2}{ e^{i\theta} - r ^2}$	$r^{ \ell }$	$1-r$
Fejér:	$\psi_h(\theta) = \frac{1}{n} \left[ \frac{\sin(n\theta/2)}{\sin(\theta/2)} \right]^2$	$(1 - \frac{ \ell }{n})^+$	$\frac{1}{n}$

$\text{supp}(\mu) = m < k$ ,  $P_k(z, \mu)$  is not uniquely defined since any polynomial with  $m$  zeros at the point mass locations will attain the minimum of zero in (1).

In [1, 2] we consider measures formed by convolving the Poisson and Fejér kernels, respectively, with the sum of point masses  $\sum_{j=1}^m \alpha_j \delta_{\theta_j}$ , where  $\delta_{\theta_j}$  is the point mass measure at  $\theta_j$  the  $\alpha_j$  are positive. Both are examples of approximate identities  $\psi_h$  with  $h \rightarrow 0$  as either  $r \rightarrow 1$  or  $n \rightarrow \infty$ , respectively, as indicated in Table 1, where  $x^+ = \max\{x, 0\}$ . It is easy to see that for any approximate identity  $\psi_h$ , we have the weak-star limit

$$\lim_{h \rightarrow 0} \psi_h * \sum_{j=1}^m \alpha_j \delta_{\theta_j} = \sum_{j=1}^m \alpha_j \delta_{\theta_j} \quad (2)$$

On the other hand, for  $m < k$ , the associated Szegő polynomials of fixed degree  $k$  do not necessarily converge. That is,  $\mu_h \rightarrow \sum_{j=1}^m \alpha_j \delta_{\theta_j}$  does not guarantee existence of the limit  $\lim_{h \rightarrow 0} P_k(z, \mu_h)$  even if  $\mu_h$  converges strongly. (See [1] for an example.)

The main point of [1] and [2] is that the  $P_k(z, \psi_h * \sum_{j=1}^m \alpha_j \delta_{\theta_j})$  do converge; moreover, the limit is the same for both kernels. We have the following

**Theorem 1.1** *Let  $\psi_h$  be either the Fejér or the Poisson kernel with the identifications in Table 1. Suppose  $\delta_{\theta_j}$  is the point mass at  $\theta = \theta_j$  with the  $\theta_j$  distinct and  $\alpha_j > 0$  for  $j = 1, 2, 3, \dots, m$ . Then*

$$\lim_{h \rightarrow 0} P_k(z, \psi_h * \sum_{j=1}^m \alpha_j \delta_{\theta_j}) = P_{k-m}(z, \nu) \prod_{j=1}^m (z - e^{i\theta_j}), \quad (3)$$

where  $\nu$  is the absolutely continuous measure with

$$\frac{d\nu}{d\theta} = \sum_{j=1}^m \prod_{p \neq j}^m \alpha_j |e^{i\theta} - e^{i\theta_p}|^2. \quad (4)$$

We seek to extend Theorem 1.1 to a larger class of kernels. We consider kernels that have moments  $\widehat{\psi}_h(\ell)$  which are convex functions of  $\ell$  for  $\ell > 0$ , or which can be expressed as a convex combination of Fejér kernels, and make a conjecture in each case. The motivation for this is that the Poisson and Fejér kernels have moments which are convex functions (though those of that latter are not strictly so). Moreover, it is possible to expand the moments of the Poisson kernel in terms of those of the Fejér. Specifically, we have

**Proposition 1.1** *Let  $\phi_n(\theta)$  denote the Fejér kernel for  $n = 1, 2, 3, \dots$ , and define  $a_{r,n} = (1-r)^2 n r^{n-1}$ . Then*

$$r^{|\ell|} = \sum_{n=1}^{\infty} a_{r,n} \widehat{\phi}_n(\ell).$$

**Proof:** We show this by writing the above as geometric series. Since  $\widehat{\phi}_n(-\ell) = \widehat{\phi}_n(\ell)$  we can assume  $\ell \geq 0$  and write

$$\begin{aligned} \sum_{n=1}^{\infty} a_{r,n} \widehat{\phi}_n(\ell) &= (1-r)^2 \sum_{n=1}^{\infty} n r^{n-1} \left(1 - \frac{\ell}{n}\right)^+ \\ &= (1-r)^2 \sum_{n=\ell+1}^{\infty} n r^{n-1} - (1-r)^2 \sum_{n=\ell+1}^{\infty} \ell r^{n-1}. \end{aligned} \quad (5)$$

Regarding the first sum in (5), we have

$$\begin{aligned} \sum_{n=\ell+1}^{\infty} n r^{n-1} &= \frac{1}{(1-r)^2} - \sum_{n=1}^{\ell} n r^{n-1} \\ &= \frac{1}{(1-r)^2} - \frac{d}{dr} \sum_{n=0}^{\ell} r^n \end{aligned}$$

$$\begin{aligned}
 &= \frac{1}{(1-r)^2} - \frac{d}{dr} \left( \frac{1-r^{\ell+1}}{1-r} \right) \\
 &= \frac{r^\ell(1+\ell-\ell r)}{(1-r)^2}.
 \end{aligned} \tag{6}$$

Regarding the second sum in (5), we have

$$\begin{aligned}
 \sum_{n=\ell+1}^{\infty} \ell r^{n-1} &= \ell \left( \sum_{n=1}^{\infty} r^{n-1} - \sum_{n=1}^{\ell} r^{n-1} \right) \\
 &= \ell \left( \frac{1}{1-r} - \frac{1-r^\ell}{1-r} \right) \\
 &= \ell \frac{r^\ell}{1-r}.
 \end{aligned} \tag{7}$$

The sum of (6) and (7) is  $\frac{r^\ell}{(1-r)^2}$ , which, with (5), proves the proposition.

An immediate consequence of Proposition 1.1 is the following.

**Corollary 1.1** *With  $\phi_n$  and  $a_{r,n}$  as in Proposition 1.1 the Poisson kernel can be expressed*

$$\psi_r(\theta) = \frac{1-r^2}{|e^{i\theta} - r|^2} = \sum_{n=1}^{\infty} a_{r,n} \phi_n(\theta)$$

## 2 Conjectures for a class of densities

Let  $f$  be a function with the following properties:

1.  $f(x) \geq 0$  for all  $x$ .
2.  $f(x) = f(-x)$ .
3.  $f$  is convex and non-increasing for  $x > 0$ .

Such functions satisfy the *Polya criterion* and thus are the characteristic functions of positive measures. (See, e.g., [3], p482.) We will call  $\{a_n\}$  and  $\psi$  a sequence and density, respectively, of *Polya type*, if  $\widehat{\phi}(n) = a_n = f(n)$  for some function  $f$  satisfying the Polya criterion. So the Poisson and Fejér kernels are densities of Polya type for  $0 < h < 1$ . We conjecture that Theorem 1.1 holds for all kernels of Polya type.

**Conjecture 1** *Let  $f_h$  be a family of Polya type functions for a continuous or discrete parameter  $h$  on  $0 < h < 1$  with  $f_h(0) = 1$  and  $\lim_{h \rightarrow 0} f_h(x) = 1$  for all  $x$ , and suppose  $\psi_h$  is the measure with  $\widehat{\psi}_h(\ell) = f_h(\ell)$ . Then (3) and holds for the kernel  $\psi_h$ , with  $\nu$  given in (4).*

A variant of Conjecture 1 is motivated by the construction of Proposition 1.1 and the fact that the Fejér kernel would seem to be a “base case” since its moments are linear rather than strictly convex. Indeed, one might suspect that any convex function of  $x$  can be expressed as a convex combination of the functions  $(1 - \frac{x}{n})^+$ .

Let  $\phi_n$  denote the Fejér kernel as in Table 1 and suppose  $\{a_n\}_{n=1}^\infty$  is a sequence of non-negative real numbers with  $\sum_{n=1}^\infty a_n = 1$ , then  $\psi(\theta) = \sum_{n=1}^\infty a_n \phi_n(\theta)$  is a density of Polya type. Now suppose that  $A_N := \{a_{N,n}\}$  is a sequence of sequences with  $\sum_{n=1}^\infty a_{N,n} = 1$  for each  $N = 1, 2, 3, \dots$  and  $\lim_{N \rightarrow \infty} a_{N,n} = 0$  for each  $n$ . If the latter holds, we write  $A_N \rightarrow 0$ . We define sequences  $A_h$  for continuous parameter  $h \rightarrow 0$  similarly, and write  $A_h \rightarrow 0$ . In either case, we simply write  $A \rightarrow 0$ , and conjecture that Theorem 1.1 holds for  $\psi_A$ .

**Conjecture 2** *Suppose  $A \rightarrow 0$  and*

$$\psi_A(\theta) = \sum_{n=1}^{\infty} a_{h,n} \phi_n(\theta).$$

*Then (1.1) holds for the kernel  $\psi_A$ ; that is,*

$$P_k(z, \psi_A * \sum_{j=1}^m \alpha_j \delta_{\theta_j}) \rightarrow P_{k-m}(z, \nu) \prod_{j=1}^m (z - e^{i\theta_j})$$

**Remarks:**

With  $a_{r,n}$  given in Corollary 1.1,  $A \rightarrow 0$  and  $\psi_{A_r}$  is the Poisson kernel, while the Fejér kernel corresponds to  $A_N = \{0, 0, 0, \dots, 1, 0, 0, \dots\}$ , with 1 in the  $N$ -th position.

We see from Table 1 that the moments of the Fejér and Poisson kernels agree to first order in  $h$ . The Fejér kernel may thus be thought of as a base case in this sense as well. The techniques in [1, 2] do not exploit this property, however. It is possible Polya-type kernels are contained in a larger class which includes those whose moments agree to first order.

## References

- [1] M. Arciero, Limits for Szegő polynomials in frequency analysis, *J. Math. Anal. Appl.* 304 (2005) 321-335.
- [2] M. Arciero, A limit theorem for Szegő Polynomials with respect to convolution of point masses with the Fejér kernel, *J. Math. Anal. Appl.*, Vol. 327, No. 2, (2007) 908-918.
- [3] W. Feller, "An introduction to probability theory and its applications" Vol II 3ed., Wiley, 1971
- [4] I.A. Geronimus, *Polynomials Orthogonal on a Circle and Interval*, New York, Pergamon Press, 1960.
- [5] W.B. Jones, O. Njåsted, Applications of Szegő Polynomials to Digital Signal Processing, *Rocky Mountain Journal of Mathematics* Vol. 21, No.1, Winter, 1991.
- [6] N. Levinson, The Wiener RMS (root mean square) error criterion in filter design and prediction, *Journal of Math. and Physics*, 25, 1947 pp. 261-268.
- [7] L. Pakula, Asymptotic zero distribution of orthogonal polynomials in sinusoidal frequency estimation, *IEEE Transactions on Information Theory*, Vol. IT 33, No.4, pp. 569-576, July 1987.
- [8] K. Pan, E.B. Saff, Asymptotics for zeros of Szego polynomials, *Journal of Approximation Theory*, Vol. 71, No. 3, pp. 239-251, Dec. 1992.
- [9] V. Petersen, Modification of a method using Szegő polynomials in frequency analysis: the V-process, *Journal of Computational and Applied Mathematics*, Volume 133, Issues 1-2 August 2001, pp. 535-544.
- [10] B. Simon, *Orthogonal Polynomial on the Unit Circle* AMS Colloquium Series, American Mathematical Society, Providence, 2005.
- [11] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, MIT Press, Cambridge, Wiley, New York, 1949.

# **A computational approach to the determination of nets**

Hans Fetter and Juan H. Arredondo R.

Departamento de Matemáticas.

Universidad Autónoma Metropolitana-Iztapalapa.

Av. San Rafael Atlixco # 186.

C. P. 09340. México, D. F.

México.

Keywords and phrases: Unfolding, dodecahedron, Shepard's conjecture.

2000 Mathematics Subject Classification: 05B30, 52B05

## **Abstract**

An unfolding of a polyhedron consists of cutting the boundary of it along some of its edges in such a way that one can flatten out the remaining set in the plane in a single piece. An unfolding is a net if it does not overlap itself. No convex polyhedra has been found which does not have a net, though almost all of its unfoldings overlap. In the particular case of the Dodecahedron, we show by means of both theoretical and computational considerations that every unfolding is a net.

## **1 Introduction**

There is an extensive study of unfoldable/foldable structures and applications are different and all very interesting. Among them, we may mention the Japanese art of paper folding and the utilization of unfoldable connected structures in aerospace. See for instance <http://www.patentgenius.com/patent/6920733.html>

Even very regular polyhedra have an overlapping unfolding, although it is possible also to construct an unfolding of the polyhedron which is a net [3]. Furthermore, Schevon [2] shows for some class of polyhedra that almost all unfoldings of a polyhedron in the class overlaps.

In the particular case of the Dodecahedron one can ask if it has an overlapping unfolding. The total number of unfoldings for the Dodecahedron is 43380. We give an answer to this question by means of theoretical and computational aids. See also [1] where a purely theoretical approach is considered.

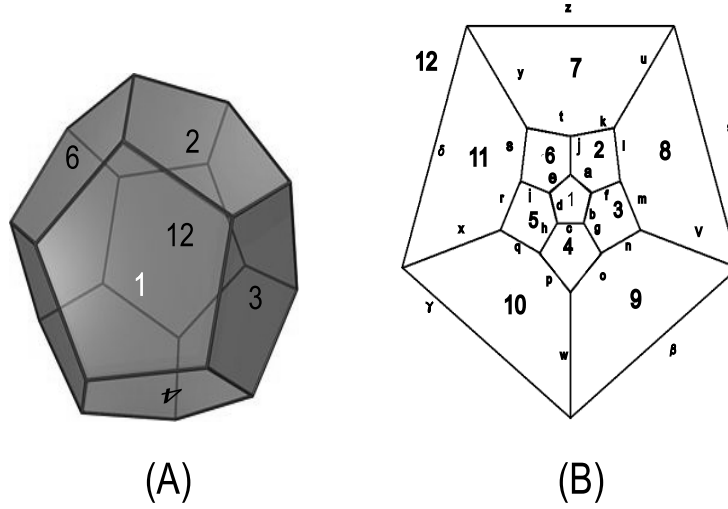


Figure 1: (A): The Platonic solid. (B): Graphic representation of the Dodecahedron.

This problem is related to Shepard's conjecture which states that any convex polyhedron has at least one net. Up to date there has not been found a convex polyhedron for which every unfolding overlaps, thus reinforcing the conjecture.

## 2 Theoretical framework

The regular dodecahedron is the Platonic solid composed of 20 vertices, 30 edges and 12 pentagonal faces. See figure 1 (A).

### Definition 2.1.

- (i) *An unfolding of a polyhedron consists of cutting the boundary of it along some of its edges in such a way that one can flatten out the remaining set in the plane in a single piece.*
- (ii) *An unfolding is a net if it does not overlap itself.*

**Definition 2.2.** *A single chain is a set of faces belonging to an unfolding of a polyhedron and such that anyone of its elements shares*

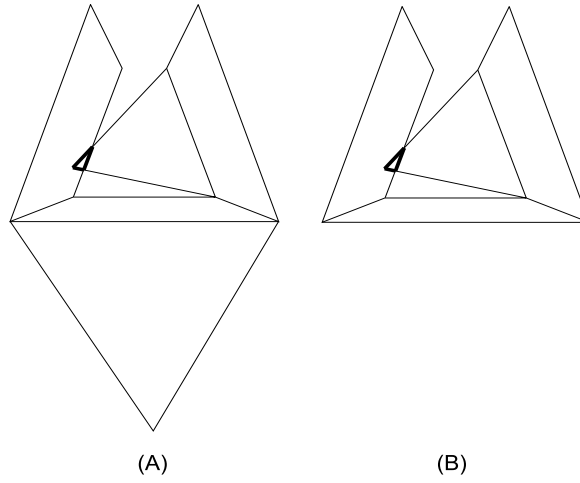


Figure 2: (A): Unfolding of a polyhedron. (B): Single chain of faces determining the overlapping. Obtained from the Wolfram Demonstrations Project.

*at least one edge and at most two edges. The number of faces in the single chain is called its length.*

Figure 2 (A) shows an unfolding of a polyhedron, which is not a net. 2 (B) exhibits the single chain determining this overlapping. In figure 3 we have an unfolding of the Dodecahedron without overlapping, and therefore it is a net.

Our strategy to solve the problem we stated is quite simple and direct. From the lemma 2.1 below, to check if every unfolding is a net one needs only to look at all single chains of lengths one to twelve. This observation reduces the problem not merely because of the number of unfoldings (43380) is greater than that of the single chains ( $\approx 5000$ ), but to construct and analyze is easier for a single chain than for an unfolding.

**Lemma 2.1.** *Every unfolding of the Dodecahedron with an overlapping has a self intersecting single chain of pentagons of length less or equal twelve.*

*Proof:* Suppose that  $\mathcal{U} = \{ f_1, \dots, f_{12} \}$  is an unfolding of the dodecahedron having an overlapping. Let  $f_k$  and  $f_\ell$  a pair of faces of



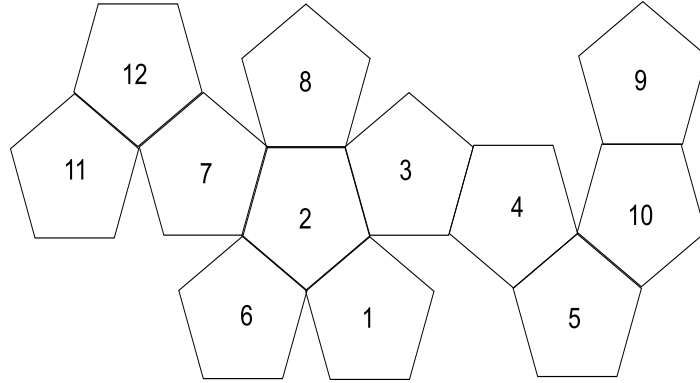


Figure 3: Unfolding of the Dodecahedron which is a net.

$\mathcal{U}$  with non void intersection and not being adjacent in the unfolding  $\mathcal{U}$ . Since every unfolding is a connected set, there must exist a subset  $\{f_k, \dots, f_\ell\}$  of  $\mathcal{U}$  forming a single chain of pentagons of length less or equal to twelve. This proves the lemma.  $\square$

**Theorem 2.1.** *If every single chain of pentagons of length less or equal to twelve does not intersect itself, then every unfolding of the Dodecahedron is a net.*

*Proof:* Since every unfolding that overlaps contains a self intersecting single chain, from the previous lemma the result follows at once.  $\square$

### 3 Main Theorem

Now we will prove that every single chain does not intersect itself. In order to show this, we use a computational algorithm that calculates and analyzes every single chain of the Dodecahedron.

**Theorem 3.1.** *Every unfolding of the Dodecahedron is a net.*

*Proof:* From theorem 2.1 we need only to check that every single chain does not intersect itself.

The strategy to do it is conceptually simple and is described in the next three steps.

(i) First step: Let  $1 \leq k \leq 12$  and  $\{f_1, \dots, f_k\}$  denote a single chain of the Dodecahedron. We note that for  $k = 1, \dots, 5$  one can check just by inspection that no single chain intersects itself.

(ii) Second step: We proceed iteratively. To check if there is a single chain of length 6 that intersects itself, we only have to verify that the pentagons at the extremes in the single chain do not intersect. Precisely, suppose that there is a single chain of length 6 that intersects itself. If  $f_i$  and  $f_j$  are not adjacent pentagons of the single chain having non void intersection, then there is a subset of pentagons  $\{f_{\ell_1}, f_{\ell_2}, \dots, f_{\ell_k}\}$  ( $2 \leq k \leq 6$ ;  $f_{\ell_1} = f_i$ ,  $f_{\ell_k} = f_j$ ) of the original single chain that forms itself another single chain, which we denote by  $S'$ . This single chain intersects itself. From the fact that no single chain of length less or equal to 5 has an overlapping,  $S'$  must have length 6. Therefore, it is the original single chain and the pentagons overlapping are the extremes of the original one. This reduces the problem to check whether the pentagons at the extremes in every single chain of length 6 do not overlap. As a consequence, if we show that the extremes of a single chain of length 6 do not intersect, then we would have proved that no such single chains have an overlapping. This argument can be applied inductively for  $k = 7, \dots, 12$ .

(iii) Third step: Construction and analysis of single chains. Our algorithm calculates systematically each single chain of lengths  $6 \leq k \leq 12$  and verifies that there is no overlapping. Therefore, the theorem is proved  $\square$

### 3.1 Description of the algorithm

We describe the algorithm. The program uses the parameter  $\ell$ =length of a single chain. The single chain is represented by a list  $S$  of  $\ell$  numbers between 1 to 12. By a previous routine we calculate all single chains of lengths  $\ell = 6, \dots, 12$ . The parameter  $D$  is a *data* matrix containing the structure of the Dodecahedron, as given in Table 1. The algorithm gives as output at a stage a list  $L$  of numbers that the computer associates to vectors connecting the centers of adjacent pentagons. See figures 4 and 9. Here we mean by the center of the

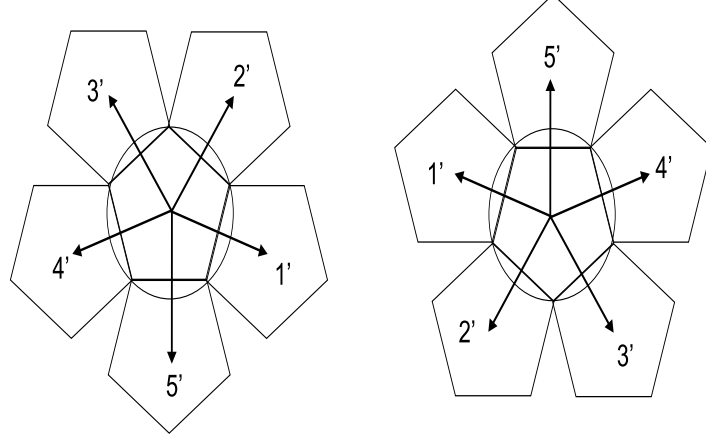


Figure 4: Directions 1', 2', 3', 4' and 5' are associated with the vectors pointing to the center of adjacent pentagons.

pentagon, the center of the circle circumscribing the pentagon. This fixes the coordinates of the center of each pentagon in the single chain. By another routine, we draw the graphic of the single chain  $S$  and determine if it intersects itself.

$a$	$b$	$c$	$d$	$e$	1	$\{2, a\}$	$\{3, b\}$	$\{4, c\}$	$\{5, d\}$	$\{6, e\}$
$a$	$j$	$k$	$l$	$f$	2	$\{1, a\}$	$\{6, j\}$	$\{7, k\}$	$\{8, l\}$	$\{3, f\}$
$m$	$n$	$g$	$b$	$f$	3	$\{8, m\}$	$\{9, n\}$	$\{4, g\}$	$\{1, b\}$	$\{2, f\}$
$g$	$o$	$p$	$h$	$c$	4	$\{1, c\}$	$\{3, g\}$	$\{9, o\}$	$\{10, p\}$	$\{5, h\}$
$d$	$h$	$q$	$r$	$i$	5	$\{1, d\}$	$\{4, h\}$	$\{10, q\}$	$\{11, r\}$	$\{6, i\}$
$e$	$i$	$s$	$t$	$j$	6	$\{1, e\}$	$\{5, i\}$	$\{11, s\}$	$\{7, t\}$	$\{2, j\}$
$k$	$t$	$y$	$z$	$u$	7	$\{2, k\}$	$\{6, t\}$	$\{11, y\}$	$\{12, z\}$	$\{8, u\}$
$l$	$u$	$\Omega$	$v$	$m$	8	$\{2, l\}$	$\{7, u\}$	$\{12, \Omega\}$	$\{9, v\}$	$\{3, m\}$
$n$	$v$	$\beta$	$w$	$o$	9	$\{3, n\}$	$\{8, v\}$	$\{12, \beta\}$	$\{10, w\}$	$\{4, o\}$
$p$	$w$	$\gamma$	$x$	$q$	10	$\{4, p\}$	$\{9, w\}$	$\{12, \gamma\}$	$\{11, x\}$	$\{5, q\}$
$r$	$x$	$\delta$	$y$	$s$	11	$\{5, r\}$	$\{10, x\}$	$\{12, \delta\}$	$\{7, y\}$	$\{6, s\}$
$z$	$\delta$	$\gamma$	$\beta$	$\Omega$	12	$\{7, z\}$	$\{8, \Omega\}$	$\{9, \beta\}$	$\{10, \gamma\}$	$\{11, \delta\}$

Table 1. Symbolic representation of the Dodecahedron.

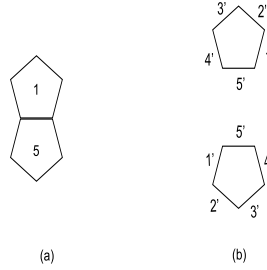


Figure 5: Figure (a) shows the orientation chosen for a single chain. Figure (b) shows the orientations given for a face “up” and a face “down”, respectively.

Note that on symmetry reasons one can assume that all single chains start at the pentagon labelled **1**. See figure 1. Furthermore, one can see by similar arguments that it is enough to analyze single chains having faces **1, 5, 4, ...** and **1, 5, 10, ...**. The algorithm calculates in a previous routine all these single chains.

We first show how the algorithm works with an example. Suppose that we want to analyze the single chain  $\{\mathbf{1}, \mathbf{5}, \mathbf{4}, \mathbf{3}, \mathbf{8}, \mathbf{12}, \mathbf{7}\}$ .

Certainly we can suppose that pentagon **1** is placed so that its edge  $d$  is horizontal. See figure 1. Then pentagon **5** has its edge  $d$  also horizontal and pentagons **1** and **5** look in an unfolding as in figure 5(a). It follows that **4** must be attached to **5** at the common edge  $h$ . See table 1 and figure 1. By simple geometry one realizes that pentagon **4** must be attached to **5** in such a way that **4** can be seen as a translation (*not a rotation*) of pentagon **1**. Similarly, pentagon **3** must be attached to **4** and this can be done by translation of pentagon **5**. In a similar way, **8** is attached to **3** by a translation of **1** and **12** is attached to **8** by a translation of **5**. By simple geometric arguments, this continues to hold every time that one attaches a new pentagon at an unfolding.

We assign to every pentagon an *interior* orientation depending on whether it is placed “up” or “down” as shown in figure 5(b).

Now we add pentagon **4** to **5**. See figure 5(a). **5** has already the orientation given by the condition that is placed “down” and its edge

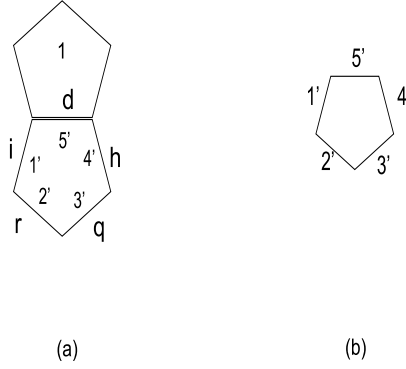


Figure 6: Orientation of face 5.

$d$  is placed horizontally. From table 1, fifth row, the first five data correspond to the edges of pentagon **5**. We know that edge  $d$  has already been assigned number  $5'$  which yields the *interior* counterclockwise orientation

$$(d, 5'), (h, 4'), (q, 3'), (r, 2'), (i, 1').$$

See figures 1 and 6.

We proceed to give an orientation to pentagon **4**. This pentagon is placed “up” and the common edge  $h$  with pentagon **5** has already been assigned  $4'$ . Looking at table 1, fourth column, the orientation obtained for pentagon **4** is

$$(g, 2'), (o, 1'), (p, 5'), (h, 4'), (c, 3').$$

See figures 1 and 7.

Now we add pentagon **3**, which must be placed “down” and the common edge with pentagon **4** is  $g$ . This edge has previously been assigned number  $2'$ . Looking at table 1, third column, the orientation for pentagon **3** is

$$(m, 4'), (n, 3'), (g, 2'), (b, 1'), (f, 5')$$

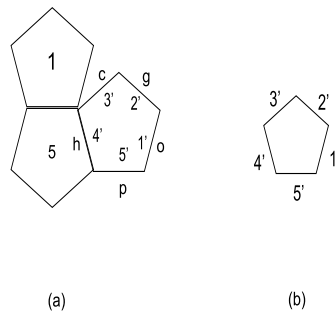


Figure 7: Orientation of face 4.

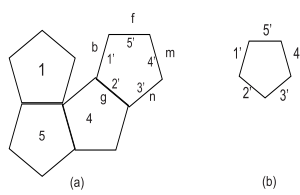


Figure 8: Orientation of face 3.

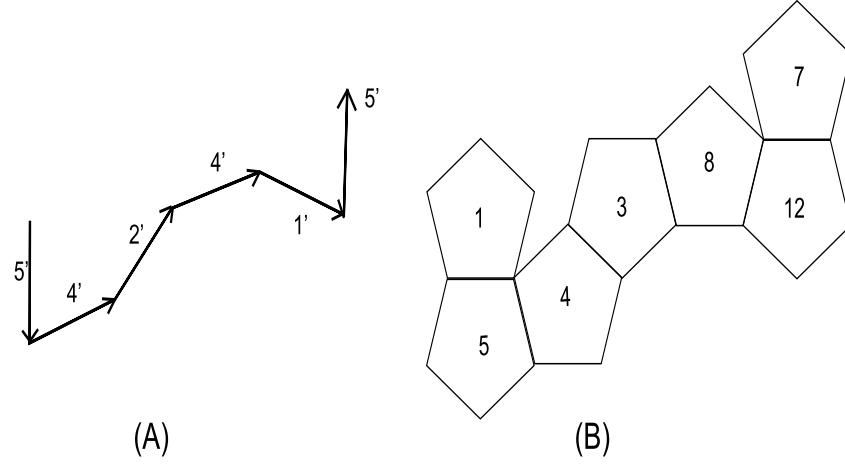


Figure 9: (A) The algorithm gives the list  $L$  associated to the directions connecting the centers of adjacent pentagons. (B) Graphic of the single chain.

See figures 1 and 8.

Initially, the center of pentagon **1** is settled so that it coincides with the origin of the chosen system of coordinates. If one continues this process, one obtains the graph of the single chain  $\{1, 5, 4, 3, 8, 12, 7\}$  as it appears in an unfolding. See figure 9 (B). To determinate if there is overlapping at this single chain, it suffices to calculate the distance between the center of pentagon **7** to the origin of coordinates. If this distance is greater than twice the radius of the circle circumscribing pentagon **1**, then there is no overlapping.

The algorithm gives as output in the case of this single chain the list  $L = \{5', 4', 2', 4', 1', 5'\}$ , which gives the vectors linking the centers of neighboring pentagons in the single chain. See figures 4 and 9.

**Algorithm 3.1:** SINGLE CHAIN ANALYSIS( $\ell, S, D$ )**Begin****comment:** Read a single chain  $S$  with elements  $S[1], \dots, S[\ell]$  $S \leftarrow \text{Read}$ **comment:** The first direction is always 5', and determines the orientation of  $S[2] = \mathbf{5}$  $L = \{5'\}$ **comment:** Orient pentagon  $S[3]$ **if**  $S[3] = 4$ **then**  $\begin{cases} \text{Pent}[3] = \{(g, 2'), (o, 1'), (p, 5'), (k, 4'), (c, 3')\} \\ \text{Append direction } 2' \text{ to } L \end{cases}$ **else if**  $S[3] = 10$ **then**  $\begin{cases} \text{Pent}[3] = \{(p, 2'), (w, 1'), (\gamma, 5'), (x, 4'), (q, 3')\} \\ \text{Append direction } 3' \text{ to } L \end{cases}$ **end if****comment:** Append directions to  $L$  of other pentagons in  $S$ **for**  $\nu_1 = 4$  **to**  $\nu_1 = \ell$ **do**  $\xi_1 = S[\nu_1]$ **comment:** Determine  $S[\nu_1] \cap S[\nu_1 - 1]$  from data matrix  $D$  $\rho_1 := S[\nu_1] \cap S[\nu_1 - 1]$ **comment:** Save direction assigned to  $\rho_1$  previously $L \leftarrow \text{Store direction assigned to } \rho_1$ **comment:** Assign orientation to pentagon  $\xi_1$  compatible

with orientation at previous face

 $\text{Pent}[\nu_1] = \{(\text{Perm}(1), 1'), (\text{Perm}(2), 2'), (\text{Perm}(3), 3'), (\text{Perm}(4), 4'), (\text{Perm}(5), 5')\}$ **end for****Return;****end program**



## References

- [1] Fetter, H. and Arredondo, J. H. *On the overlapping in the unfolding of the Dodecahedron*. Accepted for publication in Journal of Mathematicas Sciences: Advances and Applications.
- [2] Schevon, C. *Algorithms for Geodesics on Polytopes*. Ph. D. Thesis, John Hopkins University, 1989.
- [3] Schlickenrieder, W. *Nets of Polyhedra*. Ph. D. Thesis. Technische Universität Berlin. 1997.
- [4] Shephard, G. C. *Convex polytopes with convex nets*. Math. Proc. Cambridge Philos. Soc. 78 (1975), no. 3, 389–403.
- [5] Hippenmeyer, Ch. *Die Anzahl der inkongruenten ebenen Netze eines regulren Ikosaeders*, Elem. Math. 34 (1979) 61-63.
- [6] Buekenhout, F. and Parker, M. *The number of nets of the regular convex polytopes in dimension  $\leq 4$* . Discrete Math. 186 (1998), no. 1-3, 69–94.

# Kernel based Wavelets on $S^3$

S. Bernstein\*, S. Ebert†

March 15, 2009

MSC 2000: 42C40

Keywords: Wavelets, approximate identities, zonal functions

Wavelets are used to split up complicate signals into simpler parts reflecting different scales at different positions. The investigation of crystalline structures motivates studying wavelets on the three-dimensional sphere ( $S^3$ ).

## 1 Introduction

It is an interesting and though task to construct wavelets for a sphere, because there is no naive approach that works. The biggest obstacle seemed to be to define a dilation. There are several approaches and the most successful ones are those by the group of W. Freeden (see for example [8]) and those of J. P. Antoine and P. Vandergheynst ([1], [2]). The approach by J. P. Antoine and P. Vandergheynst is a group theoretical approach where as the approach by W. Freeden is done by singular integrals and zonal functions. As it is mentioned in [3] both approaches are some how equivalent, even though a deeper study of the connection of both approaches seems to be missing up to now. We are interested in wavelets on the 3 dimensional sphere, which should be motivated by the following problem. In texture analysis, i.e. the analysis of preferred crystallographic orientation, the orientation probability density function  $f$  representing the probability law of random orientations of crystal grains by volume is a major issue. In X-ray or neutron diffraction experiments spherical intensity distributions are measured which can be interpreted in terms of spherical probability distributions of distinguished crystallographic axes. In texture analysis, they are referred to as pole probability density functions. In general, if  $f$  is the orientation probability density function of a random rotation, then the spherical Radon transform  $\mathcal{R}f$

---

\*corresponding author, swanhild.bernstein@math.tu-freiberg.de

†both authors: Institute of Applied Analysis, Freiberg University of Mining and Technology, D-09599 Freiberg, Germany

integrating  $f$  over all 1-dimensional great circles  $C \subset S^3$  is simply the probability density function of  $g \in C \mid C \subset S^3$ . In this way the spherical Radon transform provides an appropriate model of the X-ray diffraction experiment of texture analysis.

To investigate the localisation properties of this spherical Radon transform we need wavelets on the 3-sphere. Our approach will use singular integrals and zonal functions.

## 2 Preliminaries

### 2.1 Surface hyperspherical harmonics

If  $\mathbb{R}^4$  is parameterized in polar co-ordinates:

$$\begin{aligned} x_1 &= r \sin \theta_3 \sin \theta_2 \cos \theta_1, \\ x_2 &= r \sin \theta_3 \sin \theta_2 \sin \theta_1, \\ x_3 &= r \sin \theta_3 \cos \theta_2, \\ x_4 &= r \cos \theta_3, \end{aligned}$$

where  $0 \leq \theta_2, \theta_3 \leq \pi$  and  $0 \leq \theta_1 \leq 2\pi$ , then the hyperspherical harmonics on  $S^3$  are defined by

$$Y_{klm} = (2i)^l l! \sqrt{\frac{2(k+l)(k-l)!}{\pi(k+l+1)!}} Y_{lm}(\theta_1, \theta_2) C_{k-1}^{l+1}(\cos \theta_3) \sin^l \theta_3,$$

where  $Y_{lm}(\theta_1, \theta_2)$  are the well-known spherical harmonics on  $S^2$ . In these co-ordinates, the (not normalized)  $SO(4)$ -invariant measure on  $S^3$  reads

$$d\sigma = \sin \theta_2 \sin^2 \theta_3 d\theta_1 d\theta_2 d\theta_3.$$

The basis  $\{Y_{klm}, 0 \leq l \leq k, -l \leq m \leq l\}$  is in fact based on the reduction of the representation of  $SO(4)$  to representations of  $SO(3)$ ; each  $Y_{klm}$  is an eigenfunction of an  $SO(3)$  subgroup of  $SO(4)$  which leaves a selected point of  $S^3$  invariant.

### 2.2 Function spaces

We have for  $1 \leq p < \infty$ , the Lebesgue space

$$L^p(S^3) = \left\{ f : \|f\|_p = \left( \frac{1}{2\pi^2} \int_{S^3} |f(x)|^p ds(x) \right)^{1/p} \right\}$$

and the weighted Lebesgue space for zonal functions, i.e. functions that depend only on the angle  $\theta$  and with  $t = \cos \theta$ :

$$L_1^p[-1, 1] = \left\{ f : \|f\|_{p,1} = \left( \frac{1}{\sqrt{\pi}} \int_{-1}^1 |f(t)|^p \sqrt{1-t^2} dt \right)^{1/p} \right\}$$

### 2.3 Funk-Hecke theorem

The ultraspherical or Gegenbauer polynomials  $\{C_n^1(t)\}$  (identical to the Chebyshev polynomials of second kind  $U_n$ ) build a complete orthonormal system on  $[-1, 1]$  with weight  $\sqrt{1-t^2}$ . We denote by  $x \cdot y = \cos \angle(x, y) =: t$  the scalar product of  $x$  and  $y$  in  $\mathbb{R}^4$ .

**Lemma 2.1.** *Let  $Y_n^l(x)$ ,  $l \in \{1, 2, \dots, (n+1)^2\}$ , be an orthonormal basis of spherical harmonics of degree  $l$ . Furthermore, let*

$$c_l(f) := \int_{S^3} f(y) \overline{Y_n^l(y)} d\sigma(y).$$

Then:

$$\sum_{l=0}^N \sum_{l=1}^{(n+1)^2} c_{l_j}(f) S_{l_j}(x)$$

tends to  $f(x)$  as  $N \rightarrow \infty$ , uniformly in  $x \in S^3$ .

*Proof.* The proof can be found in [10] or [12]. An immediate consequence is that the set of all spherical harmonics is dense in  $C(S^3)$  and hence in all spaces  $L^p(S^3)$ ,  $1 \leq p < \infty$ . It should also be mentioned a remarkable result: Given  $1 \leq p < \infty$ ,  $p \neq 2$ , there exists an  $f \in L^p(S^3)$  such that the partial sums of the Laplace series for  $f$  do not converge in the  $L^p$ -norm! (see [5] and mentioned in [10].) This is equivalent to the same result for Fourier series.

**Theorem 2.2.** *(Addition theorem) Let  $\{Y_n^l(x); l = 1, 2, \dots, (n+1)^2\}$  be the set of  $(n+1)^2$  linear independent spherical harmonics of degree  $n$  and  $\{Y_n^l(x)\}$  orthonormal on  $S^3$ , then*

$$C_n^1(x \cdot y) = \frac{2\pi^2}{n+1} \sum_{l=1}^{(n+1)^2} Y_n^l(x) Y_n^l(y), \quad x, y \in S^3.$$

A very important theorem is

**Theorem 2.3.** *(Funk-Hecke) Let  $f \in L_1^1[-1, 1]$ , then*

$$\int_{S^3} f(x \cdot y) Y_n(x) d\sigma(x) = Y_n(y) \frac{4\pi}{n+1} \int_{-1}^1 f(t) C_n^1(t) \sqrt{1-t^2} dt.$$

Especially for  $n = 0$  we obtain

$$\int_{S^3} f(x \cdot y) d\sigma(x) = 4\pi \int_{-1}^1 f(t) \sqrt{1-t^2} dt.$$

The spherical harmonics  $\{Y_n^l(x), l = 1, 2, \dots, (n+1)^2\}$  build a complete orthonormal system on  $S^3$  in  $C(S^3)$  and  $L^p(S^3)$ ,  $1 \leq p < \infty$ . Let  $f \in L^1(S^3)$  then  $f$  can be expanded into a Laplace series of spherical harmonics, we have

$$S(f; x) \sim \sum_{n=0}^{\infty} Y_n(f; x),$$

where  $Y_n(f; x)$  is given by

$$Y_n(f; x) = \frac{n+1}{2\pi^2} \int_{S^3} C_n^1(x \cdot y) f(y) d\sigma(y), \quad n = 0, 1, 2, \dots \quad (1)$$

and if  $f$  is zonal, i.e. depends only on the scalar product  $x \cdot y = t$ , we have

$$Y_n(f; t) = \hat{f}(n) C_n^1(t), \quad \text{where } \hat{f}(n) = \frac{2}{\pi} \int_{-1}^1 f(t) C_n^1(t) \sqrt{1-t^2} dt. \quad (2)$$

### 3 Spherical singular integrals

The properties of the convolutions, i.e. Young's inequalities for groups and homogeneous spaces can be found in the appendix. An important type of singular integrals are singular integrals of convolution type which are generated by a singular kernel. This technique has been introduced in the Euclidean space by Mikhlin and Pröbldorf [11], earlier by Calderon and Zygmund [6], and had been extended to spheres by Dunkl [7] and Butzer [4]. The spherical convolution is based on the convolution with zonal functions. For the sphere  $S^3$  this means following [4]

**Definition 3.1.** For  $f \in L^1(S^3)$  and  $g \in L_1^1[-1, 1]$  the convolution  $h = f * g$  of  $f$  and  $g$  is defined by

$$h(x) = \frac{1}{2\pi^2} \int_{S^3} f(y) g(x \cdot y) d\sigma(y).$$

The convolution has the following properties:

**Remark 3.2.** Let be  $f \in L^p(S^3)$  and  $g \in L_1^q$ ,  $1 \leq p, q \leq \infty$ , then is the convolution  $f * g$  is defined almost everywhere on  $S^3$  and we have Young's inequality (see for example [13]):

$$\|f * g\|_r \leq \|f\|_p \|g\|_{q,1}, \quad \frac{1}{r} = \frac{1}{q} + \frac{1}{p} - 1 \geq 0,$$

in particular

$$\|f * g\|_p \leq \|f\|_p \|g\|_{1,1} \quad \text{and} \quad \|f * g\|_q \leq \|f\|_1 \|g\|_{q,1}.$$

The Laplace series expansion of  $f * g$  has the form

$$Y_n(f * g; x) = \frac{1}{n+1} \hat{g}(n) Y_n(f; x),$$

where  $\hat{g}(n)$  and  $Y_n(f; x)$  are given by (2) and (1) respectively.

Based on this spherical convolution we are now able to define *singular integrals on the unit sphere*.

**Definition 3.3.** Let  $K_h \in L_1^1[-1, 1]$ , with  $h \in (0, 1)$ , be a family of kernels such that the coefficients satisfy

$$\hat{K}_h(0) = \frac{2}{\pi} \int_{-1}^1 K_h(t) C_0^1(t) \sqrt{1-t^2} dt = 1.$$

Then the family

$$I_h(f) = K_h * f = \frac{1}{2\pi^2} \int_{S^3} f(y) K_h(\cdot \cdot y) d\sigma(y)$$

is called a spherical singular integral while the family  $K_h$ ,  $0 \leq h < 1$ , is called its kernel.

A singular integral  $I_h$  is said to be an approximate identity in  $L^p(S^3)$ ,  $1 \leq p < \infty$ , if

$$\lim_{h \rightarrow 1^-} \|I_h f - f\|_p = 0.$$

*Remark 3.4.* In [4] the convolution integrals defined above called spherical singular integrals.

**Theorem 3.5.** Let  $K_h \in L_1^1[-1, 1]$ , with  $h \in (0, 1)$  be a family of kernels such that

1. it is a kernel of a spherical convolution integral  $I_h$ , that is, it satisfy

$$\hat{K}_h(0) = \frac{1}{2\pi^2} \int_{-1}^1 K_h(t) C_0^1(t) \sqrt{1-t^2} dt = 1;$$

2. there exists a constant  $M \geq 1$  such that for all  $h \in (0, 1)$ , we have

$$\frac{1}{2\pi^2} \int_{S^3} |K_h(t)| \sqrt{1-t^2} dt \leq M;$$

3. for every fixed  $\delta > 0$ ,

$$\lim_{h \rightarrow 1^-} \sup_{-1 \leq t \leq 1-\delta} |K_h(t)| = 0.$$

Then for every  $f \in L^p(S^3)$ ,  $1 \leq p < \infty$ , the convolution integral

$$I_h(f) = K_h * f = \frac{1}{2\pi^2} \int_{S^3} f(y) K_h(\cdot, y) d\sigma(y)$$

fulfills

**a)**  $\|I_h f\|_p \leq M \|f\|_p;$

**b)**  $\lim_{h \rightarrow 1^-} \|I_h f - f\|_p = 0.$

*Proof.* Proposition **a)** is a consequence of Young's inequality and (2.). To prove the second proposition, we define

$$S(x, \delta) = \{y \in S^3 : x \cdot y \leq 1 - \delta, \delta > 0\}.$$

Then

$$\begin{aligned} I_h(f, x) - f(x) &= \frac{1}{2\pi^2} \int_{S^3 \setminus S(x, \delta)} K_h(x \cdot y) [f(y) - f(x)] d\sigma(y) \\ &\quad + \frac{1}{2\pi^2} \int_{S(x, \delta)} K_h(x \cdot y) [f(y) - f(x)] d\sigma(y) \\ &= I_1 + I_2. \end{aligned}$$

For  $I_1$ , we use the fact

$$\int_{S^{n-1}} |K_h(x \cdot y)| d\sigma(y) = 2\pi \int_{-1}^1 |K_h(t)| (1 - t^2) dt.$$

By Young's inequality, Hölder-Minkowski's inequality and (3.) we have that a given  $\epsilon > 0$ , there exists  $h_0(\epsilon)$  such that

$$\begin{aligned} \|I_1\|_p &\leq \frac{2\pi}{2\pi^2} 2\|f\|_p \int_{-1}^1 |K_h(t)| \sqrt{1 - t^2} dt \\ &\leq \frac{2}{\pi} \|f\|_p \sup_{\{-1 \leq t \leq 1 - \delta\}} |K_h(t)| < \epsilon, \end{aligned}$$

for  $h > h_0(\epsilon)$ .

Further, we get

$$\|I_2\|_p \leq \frac{1}{2\pi^2} \int_{S(x, \delta)} |K_h(x \cdot y)| \|f(y) - f(x)\| d\sigma(y) \leq \epsilon \cdot M,$$

if  $h < 1 - \delta$ , due to Kolmogorov's compactness theorem (see for example [14] for  $L^p$ , the one-point set  $\{f\}$  being relatively compact. Hence,

$$\|I_h f - f\|_p \leq \|I_1\|_p + \|I_2\|_p < (M + 1)\epsilon,$$

for  $h < h_0(\epsilon)$ . □

**Lemma 3.6.** Assume that the kernel  $\{K_h\}_{0 < h < 1}$  is uniformly bounded in the  $L_1^1$ -norm, i.e. there exists a positive constant  $M$  (independent of  $h$ ) such that

$$\int_{-1}^1 |K_h(t)| \sqrt{1 - t^2} dt \leq M \tag{3}$$

for all  $0 < h < 1$ . Then the corresponding convolution integral  $\{I_h\}$  is an approximate identity in  $L^p(S^3)$ ,  $1 \leq p < \infty$ , if and only if

$$\lim_{h \rightarrow 1-} \hat{K}_h(n) = n + 1 \quad (4)$$

for all non-negative integers  $n$ .

*Proof.* For  $f \in C(S^3)$  we have the Laplace expansion

$$f(x) = \sum_{n=0}^{\infty} Y_n(f, x)$$

and we know that

$$(f * K_h)(x) \sim \sum_{n=0}^{\infty} \frac{1}{n+1} \hat{K}_h(n) Y_n(f, x)$$

Let  $P_l$  be a spherical polynomial, e.g. a finite linear combination of spherical harmonics. From (4) we get that  $I_h P_l$  converges to  $P_l$  for any spherical polynomial  $P_l(x)$  in the  $L^p$ -norm because in this case the Laplace series of  $I_h P_k$  has only finite many summands. Because the set of all spherical polynomials is dense in  $L^p(S^3)$  (see Lemma 2.1) the convergency in the  $L^p$ -norm follows from the Hahn-Banach theorem (density argument).

Typical examples are the Abel-Poisson integral and the Gauß-Weierstraß integral.

## 4 Continuous Wavelet Transform

### 4.1 Linear Theory

We define spherical wavelets of order  $m = -1$ .

**Definition 4.1.** Assume that  $\mu : [0, \infty) \rightarrow \mathbb{R}_+$  is a positive weight function. Let  $\{\Psi_\rho, \rho \in (0, \infty)\}$ , be a subfamily of  $L^2_1[-1, 1]$  such that the following admissibility conditions are satisfied:

- for  $n = 0, 1, \dots$

$$\int_0^\infty \hat{\Psi}_\rho(n) \mu(\rho) d\rho = n + 1, \quad (5)$$

- for all  $R \in (0, \infty)$

$$\int_{-1}^1 \left| \int_R^\infty \Psi_\rho(t) \mu(\rho) d\rho \right| \sqrt{1-t^2} dt \leq T, \quad (6)$$

where  $T$  is a positive constant independent of  $R$ .



Then  $\{\Psi_\rho\}$  is called a spherical (linear) wavelet of order  $-1$ . The function  $\Psi = \Psi_1$  is called a spherical mother wavelet. The associated (linear) wavelet transform is defined by

$$(WT)(F)(\rho, y) := \frac{1}{2\pi^2} \int_{S^3} \Psi_\rho(x \cdot y) F(x) d\sigma(x)$$

for all  $F \in L^2(S^3)$ .

Then

**Theorem 4.2.** (*Reconstruction formula*). *Let  $\{\Psi_\rho\}$  be a wavelet (of order  $-1$ ). Then  $F \in L^2(S^3)$  can be reconstructed in  $L^2$ -sense by*

$$F(y) = \int_0^\infty (WT)(\rho, y) \mu(\rho) d\rho = \int_0^\infty \frac{1}{2\pi^2} \int_{S^3} \Psi_\rho(x \cdot y) F(x) d\sigma(x) \mu(\rho) d\rho.$$

Proof: Let  $R$  be a positive number. Then

$$\begin{aligned} \int_R^\infty (WT)(F)(\rho, y) \mu(\rho) d\rho &= \frac{1}{2\pi^2} \int_R^\infty \int_{S^3} \Psi_\rho(x \cdot y) F(x) d\sigma(x) \mu(\rho) d\rho \\ &= \frac{1}{2\pi^2} \int_{S^3} \Phi_R(x \cdot y) F(x) d\sigma(x) = (\Phi_R * F)(y), \end{aligned}$$

where

$$\begin{aligned} \frac{1}{2\pi^2} \int_R^\infty \int_{S^3} \Psi_\rho(x \cdot y) F(x) d\sigma(x) \mu(\rho) d\rho &= \int_R^\infty (\Psi_\rho * F)(y) \mu(\rho) d\rho \\ &= \int_R^\infty \sum_{n=0}^\infty \frac{1}{n+1} \hat{\Psi}_\rho(n) Y_n(F; y) \mu(\rho) d\rho \\ &= \frac{1}{2\pi^2} \int_R^\infty \sum_{n=0}^\infty \hat{\Psi}_\rho(n) \int_{S^3} C_n^1(x \cdot y) F(x) d\sigma(x) \mu(\rho) d\rho \end{aligned}$$

because of (3) and (1). Hence

$$\Phi_R(x \cdot y) = \int_R^\infty \sum_{n=0}^\infty \hat{\Psi}_\rho(n) \mu(\rho) d\rho C_n^1(x \cdot y).$$

According to our construction the kernels  $\Phi_R \in L_1^2[-1, 1]$  for all  $R > 0$  (cf. [4]) and due to (5)

$$\lim_{R \rightarrow 0+} \hat{\Phi}_R(n) = n + 1 \quad \forall n \in \mathbb{N}_0.$$

From

$$\Phi_R(t) = \int_R^\infty \sum_{n=0}^\infty \hat{\Psi}_\rho(n) C_n^1(t) \mu(\rho) d\rho = \int_R^\infty \Psi_\rho(t) \mu(\rho) d\rho$$

and (6) we deduce that the kernel  $\{\Phi_R\}$  is uniformly bounded in the sense of (3) and thus

$$\lim_{R \rightarrow 0+} \frac{1}{2\pi^2} \int_{S^3} \Phi_R(x \cdot y) F(y) d\sigma(y) = \sum_{n=0}^{\infty} \frac{n+1}{2\pi^2} Y_n(F; x) = F(x). \quad \square$$

*Remark 4.3.* Analogously to [8] there would be another condition:  
for  $R \in (0, \infty)$

$$\sum_{n=0}^{\infty} (n+1) \int_R^{\infty} \hat{\Psi}(n) \mu(\rho) d\rho < \infty.$$

This condition is needed for the non-negativity of the kernel which is important for the approximation properties but neither for the definition of wavelets nor the reconstruction formula.

*Remark 4.4.* The defining properties of a spherical wavelet of order  $-1$  do not presume the zero-mean property of  $\Psi_\rho$  (or equivalently  $\hat{\Psi}_\rho(0) = 0$ ). This coincides with the observation in [8] for the sphere  $S^2$  and in [1] that the zero-mean condition is not a necessary condition. Here, we have a substantial difference to the classical (Euclidean) wavelet concept on the real line.

**Definition 4.5.** Let  $\{\Psi_\rho\}$ ,  $\rho \in (0, \infty)$ , be a subfamily of  $L_1^2[-1, 1]$  such that the admissibility conditions are satisfied:

- for  $n = 1, 2, \dots$

$$\int_0^{\infty} \hat{\Psi}_\rho(n) \mu(\rho) d\rho = n + 1, \quad (7)$$

- for  $\rho \in (0, \infty)$

$$\hat{\Psi}_\rho(0) = 0,$$

- for all  $R \in (0, \infty)$

$$\int_{-1}^1 \left| \int_R^{\infty} \Psi_\rho(t) \mu(\rho) d\rho \right| \sqrt{1-t^2} dt \leq T, \quad (8)$$

where  $T$  is a positive constant independent of  $R$ .

Then  $\{\Psi_\rho\}$  is called a spherical (linear) wavelet of order 0. The function  $\Psi = \Psi_1$  is called a spherical mother wavelet and the associated wavelet transform is defined by

$$(WT)(F)(\rho; y) = (\Psi_\rho(y), F)_{L^2(S^3)}$$

for all  $F \in L^2(S^3)$ .

**Theorem 4.6.** *Let  $\{\Psi_\rho\}$  be a wavelet of order 0. Suppose that  $F \in L^2$  satisfies  $Y_0(F; x) = 0$ . Then*

$$F(x) = \frac{1}{2\pi^2} \int_0^\infty (WT)(F)(\rho, x) \mu(\rho) d\rho.$$

Proof: Let  $R$  be a positive number. Then

$$\begin{aligned} \int_R^\infty (WT)(F)(\rho, y) \mu(\rho) d\rho &= \frac{1}{2\pi^2} \int_R^\infty \int_{S^3} \Psi_\rho(x \cdot y) F(x) d\sigma(x) \mu(\rho) d\rho \\ &= \int_{S^3} \Phi_R(x \cdot y) F(x) d\sigma(x) = (\Phi_R * F)(y), \end{aligned}$$

where

$$\begin{aligned} \frac{1}{2\pi^2} \int_R^\infty \int_{S^3} \Psi_\rho(x \cdot y) F(x) d\sigma(x) \mu(\rho) d\rho &= \int_R^\infty (\Psi_\rho * F)(y) \mu(\rho) d\rho \\ &= \int_R^\infty \sum_{n=0}^\infty \frac{1}{n+1} \hat{\Psi}_\rho(n) Y_n(F; y) \mu(\rho) d\rho \\ &= \frac{1}{2\pi^2} \int_R^\infty \sum_{n=1}^\infty \hat{\Psi}_\rho(n) \int_{S^3} C_n^1(x \cdot y) F(x) d\sigma(x) \mu(\rho) d\rho \end{aligned}$$

because of (3) and (1). Hence we can choose

$$\Phi_R(x \cdot y) = 1 + \int_R^\infty \sum_{n=1}^\infty \hat{\Psi}_\rho(n) \mu(\rho) d\rho C_n^1(x \cdot y).$$

According to our construction the kernels  $\Phi_R \in L_1^2[-1, 1]$  for all  $R > 0$  (cf. [4]) and due to (7)

$$\lim_{R \rightarrow 0+} \hat{\Phi}_R(n) = n+1 \quad \forall n \in \mathbb{N}_0.$$

From

$$\Phi_R(t) = 1 + \int_R^\infty \sum_{n=1}^\infty \hat{\Psi}_\rho(n) C_n^1(t) \mu(\rho) d\rho = 1 + \int_R^\infty \Psi_\rho(t) \mu(\rho) d\rho$$

and (8) we deduce that the kernel  $\{\Phi_R\}$  is uniformly bounded in the sense of (3) and thus

$$\lim_{R \rightarrow 0+} \frac{1}{2\pi^2} \int_{S^3} \Phi_R(x \cdot y) F(y) d\sigma(y) = \sum_{n=1}^\infty \frac{n+1}{2\pi^2} Y_n(F; x) = F(x). \quad \square$$

*Remark 4.7.* The linear analysis here may be formally interpreted as bilinear analysis:

$$F(x) = \int_0^\infty \int_{S^3} (WT)(F)(\rho; y) \delta(y \cdot x) d\sigma(y) \mu(\rho) d\rho,$$

where

$$\delta(x \cdot y) = \sum_{n=0}^\infty (n+1) C_n^1(x \cdot y)$$

is the Dirac distribution.

## 4.2 Bilinear Theory

The concept of dilation and translation becomes more evident in bilinear theory.

**Definition 4.8.** Let  $\mu : [0, \infty) \rightarrow \mathbb{R}_+$  be a positive weight function and  $\{\Psi_\rho, \rho \in (0, \infty)\}$ , be a subfamily of  $L^2_1[-1, 1]$  such that the following admissibility conditions are satisfied:

- for  $n = m, m + 1, \dots$

$$\int_0^\infty \hat{\Psi}_\rho(n) \mu(\rho) d\rho = (n + 1)^2, \quad (9)$$

- for  $n = 0, 1, \dots, m$ , and all  $\rho \in (0, \infty)$

$$\hat{\Psi}_\rho(n) = 0, \quad (10)$$

- for all  $R \in (0, \infty)$

$$\int_{-1}^1 \left| \int_R^\infty (\Psi_\rho * \Psi_\rho)(t) \mu(\rho) d\rho \right| \sqrt{1 - t^2} dt \leq T, \quad (11)$$

where  $T$  is a positive constant independent of  $R$ .

Then  $\{\Psi_\rho\}$  is called a spherical wavelet of order  $m$ . The function  $\Psi = \Psi_1$  is called a spherical mother wavelet. The associated (linear) wavelet transform is defined by

$$(WT)(F)(\rho, y) := \frac{1}{2\pi^2} \int_{S^3} \Psi_\rho(x \cdot y) F(x) d\sigma(x)$$

for all  $F \in L^2(S^3)$ .

Then

**Theorem 4.9.** (*Reconstruction formula*). Let  $\{\Psi_\rho\}$  be a wavelet (of order  $m$ ). The wavelet transform is invertible on the range of all functions  $F \in L^2(S^3)$  with  $\hat{F}(k, i) = 0$  for  $k = 0, 1, \dots, m$  and  $i = 1, \dots, (k + 1)^2$ , in  $L^2$ -sense by

$$F(y) = \frac{1}{2\pi^2} \int_{S^3} \int_0^\infty (WT)(\rho, y) \Psi_\rho(x \cdot y) \mu(\rho) d\rho d\sigma(x).$$

Proof: Let  $R$  be a positive number. Then

$$\begin{aligned}
& \frac{1}{2\pi^2} \int_{S^3} \int_0^\infty (WT)(\rho, y) \Psi_\rho(x \cdot y) \mu(\rho) d\rho d\sigma(x) \\
&= \frac{1}{2\pi^2} \int_{S^3} \int_0^\infty \frac{1}{2\pi^2} \int_{S^3} \Psi_\rho(x \cdot z) F(z) d\sigma(z) \Psi_\rho(x \cdot y) \mu(\rho) d\rho d\sigma(x) \\
&= \frac{1}{2\pi^2} \int_{S^3} \int_0^\infty \underbrace{\frac{1}{2\pi^2} \int_{S^3} \Psi_\rho(x \cdot z) \Psi_\rho(x \cdot y) \mu(\rho) d\rho d\sigma(x)}_{\Theta(y \cdot z)} F(z) d\sigma(z) \\
&= \frac{1}{2\pi^2} \int_{S^3} \int_0^\infty \sum_{k=0}^\infty \hat{\Psi}_\rho^2(k) \int_{S^3} C_k^1(x \cdot z) C_k^1(x \cdot y) \mu(\rho) d\rho d\sigma(x) F(z) d\sigma(z) \\
&= \frac{1}{2\pi^2} \int_{S^3} \sum_{k=0}^\infty \frac{1}{k+1} \int_R^\infty \hat{\Psi}_\rho^2 m(\rho) d\rho C_k^1(x \cdot z) F(z) d\sigma(z) \\
&= (\Theta_R * F)(y).
\end{aligned}$$

Hence

$$\Theta(y \cdot z) = \sum_{k=m}^\infty \frac{1}{k+1} \int_R^\infty \hat{\Psi}_\rho^2 \mu(\rho) d\rho C_k^1(x \cdot z)$$

and uniformly bounded due to (11), further because of (9) we have

$$\lim_{R \rightarrow 0+} \hat{\Theta}_R(k) = k+1, \quad \text{for all } k = m+1, m+2, \dots \quad (12)$$

Hence by (11) and (12)  $\{\Theta_R, R > 0\}$  is the kernel of an approximate identity and thus

$$\lim_{R \rightarrow 0} (\Theta_R * F)(y) = F(y)$$

in the  $L^2(S^3)$ -norm.  $\square$

We can rewrite the reconstruction formula as:

$$F = \int_0^\infty (\Psi_\rho * \Psi_\rho * F)(\cdot) \mu(\rho) d\rho.$$

In the bilinear theory reasonable to introduce a scaling-function.

**Definition 4.10.** The corresponding scaling-function  $\{\Phi_R, R > 0\}$  for a family of wavelets  $\{\Psi_\rho, \rho > 0\}$  of order  $m$  is defined by

$$\hat{\Phi}_R(k) := \begin{cases} k+1, & k = 0, 1, \dots, m, \\ \left( \int_R^\infty \hat{\Psi}_\rho^2(k) \mu(\rho) d\rho \right)^{\frac{1}{2}}, & k = m+1, \dots \end{cases}$$

It can be shown that that the scaling function  $\Phi_R$  belongs to  $L^2(S^3)$  and that

$$\lim_{R \rightarrow 0} (\Phi_R * \Phi_R * F)(y) = \lim_{R \rightarrow 0} (\Theta_R * F)(y) = F(y).$$

For details see [9].

## 5 Wavelets

The wavelets of order  $m$  corresponding to the Gau-Weierstra kernel look as follows:

$$\hat{\Psi}_R = \begin{cases} 0, & k = 0, 1, \dots, m, \\ (2k(k+1)^2(k+2)e^{-2k(k+2)R}\mu^{-1}(R))^{\frac{1}{2}}, & k = m+1, m+2, \dots \end{cases}$$

Those corresponding to the Abel-Poisson kernel are

$$\hat{\Psi}_R = \begin{cases} 0, & k = 0, 1, \dots, m, \\ (2k(k+1)^2e^{-2kR}\mu^{-1}(R))^{\frac{1}{2}}, & k = m+1, m+2, \dots \end{cases}$$

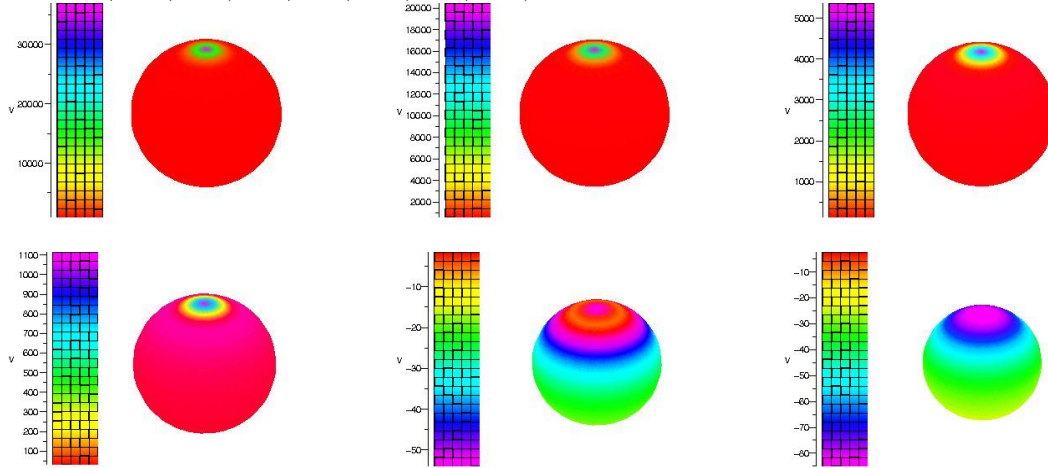
We plot Abel-Poisson wavelets of order 0 and choose  $\mu(\rho) = \frac{1}{\rho^4}$ . It is not so simple to visualize wavelets on the 3D sphere, which is a subset of  $\mathbb{R}^4$ . To overcome the dimensional problem we consider only the upper half sphere

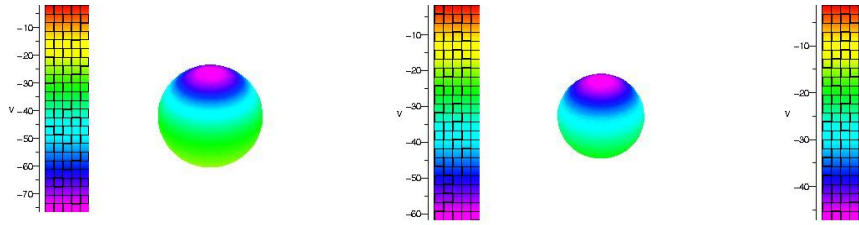
$$S_+^3 := \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : \sum_{i=1}^4 x_i^2 = 1, \text{ and } x_1 \geq 0\},$$

which can be identified with the unit ball in  $\mathbb{R}^3$  by

$$\sum_{i=1}^4 x_i^2 = 1 \iff \sum_{i=2}^4 x_i^2 = 1 - x_1^2 \leq 1$$

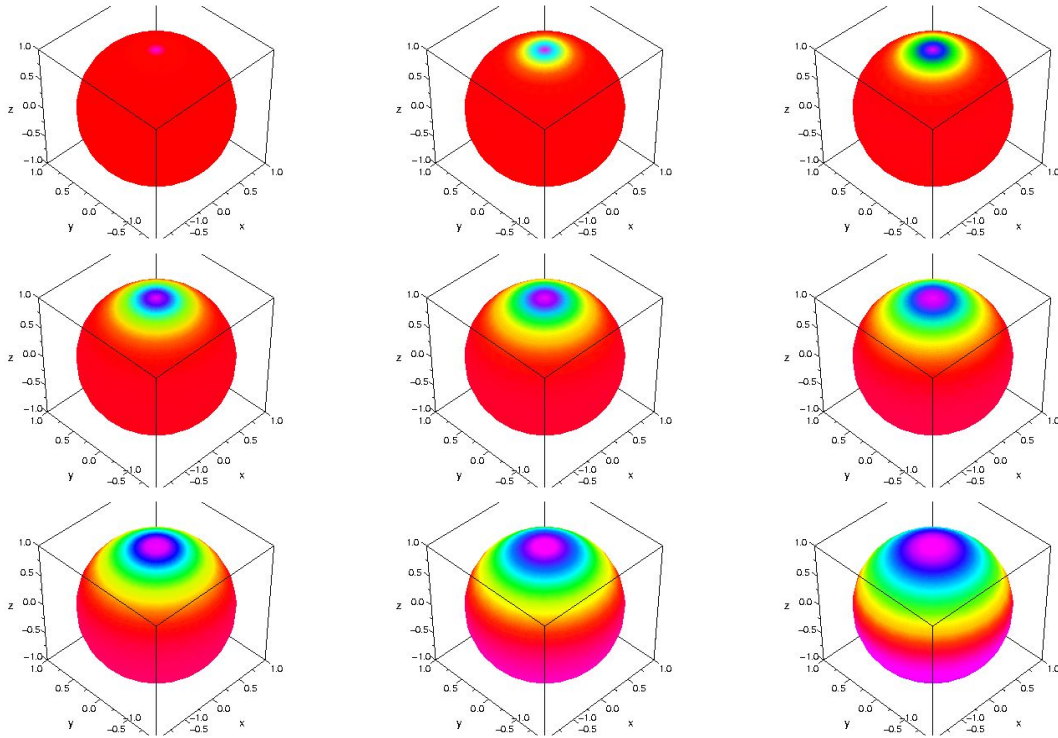
For a fixed  $x$  we obtain a 2D sphere in  $\mathbb{R}^3$  with radius  $\sqrt{1-x^2}$ . The union of all spheres with  $0 \leq x_1 \leq 1$  is the unit ball in  $\mathbb{R}^3$ . Abel-Poisson wavelet with  $\rho = 0.2$  on the slices  $x = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ :





It is easily seen that the wavelets localization is good, small positive and negative values are due to the fact that  $\rho = 0.2$ .

In the next figures the influence of  $\rho$  is demonstrated by the slice  $x_1 = 0$  for varying values of  $\rho$ . The delta peak is very good for small  $\rho$  and expands for larger values of  $\rho$ . Abel-Poisson wavelet with  $\rho = 0 \dots 3$  on the slice  $x = 0$



## References

- [1] J.-P. Antoine and P. Vandergheynst, Wavelets on the 2-sphere: A group-theoretical approach, *Appl. Comput. Harmon. Anal.* **7**, 262–291, (1999).
- [2] J.-P. Antoine and P. Vandergheynst, Wavelets on the n-sphere and other manifolds, *J. Math. Physics* **39**, 3987–4008, (1998).
- [3] J.-P. Antoine, L. Demanet, L. Jacques and P. Vandergheynst, Wavelets on the Sphere:

- Implementation and Approximations, *Appl. Comput. Harmon. Anal.* **13**, No.3, 177-200 (2002).
- [4] H. Berens, P. L. Butzer and S. Pawelke, Limitierungsverfahren von Reihen mehrdimensionaler Kugelfunktionen und deren Saturierungsverfahren, *Publ. of the Research Institute for Mathematical Sciences, Kyoto Univ. Ser. A*, **4**, 1968, pp. 201–268.
  - [5] A. Bonami and J.-L. Clerc, *Sommes de Cesàro et multiplicateurs des développements en harmoniques sphériques*, *Trans. Amer. Math. Soc.*, **138**, (1973), 223-263.
  - [6] A. P. Calderón and A. Zygmund, On a problem of Mihlin, *Trans. Amer. Math. Soc.* **78**, 209–224 (1955).
  - [7] C. F. Dunkl, Operators and harmonic analysis on the sphere, *Trans. Amer. Math. Soc.* **125**, 250–263 (1966).
  - [8] W. Freeden, T. Gervens and M. Schreiner, *Constructive Approximation on the Sphere with Applications to Geomathematics*, Numerical Mathematics and Scientific Computation, Oxford Scienes Publ., Clarendon Press, Oxford, 1998.
  - [9] Ebert, S., Wavelets on the three-dimensional sphere, Diplomarbeit (diploma thesis), TU Bergakademie Freiberg, 2008.
  - [10] H. Kalf, *On the Expansion of a Function in Terms of Spherical Harmonics in Arbitrary Dimensions*, *Bull. Belg. Math. Soc.* **2** (1995), 361-380.
  - [11] S. G. Mikhlin and S. Prößdorf, *Singular Integral Operators*, Springer-Verlag Berlin, 1986.
  - [12] K. Müller, *Analysis of Spherical Symmetries in Euclidean Spaces*, *Appl. Math. Sciences* **129**, Springer-Verlag New York, 1998.
  - [13] N. R. Wallach, *Harmonic analysis on homogeneous spaces*, Marcel Dekker Inc., New York, Pure and Applied Mathematics, No. 19, 1973.
  - [14] J. Wloka, *Partielle Differentialgleichungen*, BSB B.G. Teubner Verlagsgesellschaft, Leipzig, 1982.



# A Taste of Ideal Projectors

Boris Shekhtman  
 Department of Mathematics and Statistics  
 University of South Florida,  
 Tampa, FL 33620  
 boris@math.usf.edu  
<http://shell.cas.usf.edu/~boris/>

March 16, 2009

## Abstract

We survey the properties of ideal projectors and structure of the family of ideal projectors onto a given finite-dimensional space of polynomials. Particularly we establish relations between ideal projectors, commuting matrices, zero-dimensional ideals, solutions of systems of PDEs and certain topics in algebraic geometry, such as Hilbert and border schemes.

## 1 Menu

As the title indicates, this survey does not offer a full course on ideal interpolation but rather, following the culinary analogy, a sample of what is available in this exiting area of multivariate interpolation. Although many of the results presented here are true for the real field as well as the complex field, I will limit myself to working in the space  $\mathbb{C}[\mathbf{x}] := \mathbb{C}[x_1, \dots, x_d]$  of polynomials in  $d$  variables with complex coefficients.

**Definition 1.1** (*[Bi]*) *A linear idempotent operator  $P : \mathbb{C}[\mathbf{x}] \rightarrow \mathbb{C}[\mathbf{x}]$  is called an ideal projector if  $\ker P$  is an ideal in  $\mathbb{C}[\mathbf{x}]$ .*

Lagrange interpolation projectors, Taylor projectors and, in one variable, Hermite interpolation projectors are all examples of ideal projectors. For this reason the ideal interpolation holds a promise of elegant theory of multivariate extensions of univariate properties, which as a rule tend to be a messy subject. The brilliance of Birkhoff's idea is in restricting the domain of the projectors to the space (ring, algebra) of polynomials  $\mathbb{C}[\mathbf{x}]$  thus allowing a whole slue of various mathematics to come into play. As luck would have it, the problems in ideal interpolations are closely related to problems in commutative and linear algebra, algebraic geometry and PDEs. Here is what on the menu:

### 1.1 Approximation Theory (AT)

The main objective of this paper is to study ideal projector onto a fixed  $N$ -dimensional subspace  $G \subset \mathbb{C}[\mathbf{x}]$ . We will denote the family of all such projectors by  $\mathfrak{P}_G$ . By the end of this paper we will get to a remarkable fact: the structure (geometric, metric, topological) of  $\mathfrak{P}_G$  depends on  $G$  and differs for different spaces  $G$  of the same dimension.

### 1.2 Commutative Algebra (CA)

Every ideal projector  $P \in \mathfrak{P}_G$  determines a decomposition

$$\mathbb{C}[\mathbf{x}] = J \oplus G \quad (1.1)$$

with the ideal  $J = \ker P$ . Thus studying  $\mathfrak{P}_G$  is equivalent to studying the set  $\mathfrak{J}_G$  of ideals in  $\mathbb{C}[\mathbf{x}]$  that complement  $G$ . Given an ideal  $J \in \mathfrak{J}_G$ , the space  $G$  spans the finite-dimensional quotient algebra  $\mathbb{C}[\mathbf{x}]/J$ .

### 1.3 Algebraic Geometry (AG)

Every ideal  $J \in \mathbb{C}[\mathbf{x}]$  generates a subset

$$\mathcal{Z}(J) = \{\mathbf{z} \in \mathbb{C}^d : f(\mathbf{z}) = 0 \text{ for all } f \in J\} \quad (1.2)$$

which is called an affine algebraic set (variety, scheme). Given a basis  $\mathbf{g} = (g_1, \dots, g_N)$  of  $G$  we will construct an affine algebraic set, called *the border scheme*  $B_{\mathbf{g}}$  that parametrizes  $\mathfrak{P}_G$  (equivalently  $\mathfrak{J}_G$ ) in a natural (continuous way).

### 1.4 Linear Algebra (LA)

A sequence of  $\mathbf{L} = (L_1, \dots, L_d)$  of commuting linear operators on  $G$  is called cyclic if there exists a vector  $g_0 \in G$  such that

$$\{f(L_1, \dots, L_d)g_0 : f \in \mathbb{C}[\mathbf{x}]\} = G \quad (1.3)$$

The vector  $g_0$  is called a cyclic vector for  $\mathbf{L}$ . Let  $\mathfrak{L}_G$  stand for the family of all such sequences. With every  $P \in \mathfrak{P}_G$  ( $J \in \mathfrak{J}_G$ ) we will associate an  $\mathbf{L} \in \mathfrak{L}_G$  and vice versa.

### 1.5 Duality (PDE)

Let  $\mathbf{g} = (g_1, \dots, g_N)$  be the linear basis for  $G \subset \mathbb{C}[\mathbf{x}]$ . Every finite-dimensional projector  $P$  on  $\mathbb{C}[\mathbf{x}]$ , ideal or not, can be written as

$$P = \sum g_k \otimes F_k \quad (1.4)$$

where  $(F_k, k = 1 : N)$  consists of functionals in  $\mathbb{C}'[\mathbf{x}]$  dual to  $(g_k, k = 1 : N)$ , i.e.,  $\langle F, g_j \rangle = \delta_{j,k}$ . The space

$$\text{span } \{F_k\} = \text{ran } P^* = (\ker P)^\perp \quad (1.5)$$

is correct for  $G$ , thus for every  $F \in \text{ran } P^*$  we have  $F(f) = F(Pf)$  for all  $f \in \mathbb{C}[\mathbf{x}]$ . In other words  $P$  “interpolates” the functionals  $F \in (\ker P)^\perp$ . We will identify the functionals in  $\mathbb{C}'[\mathbf{x}]$  with formal power series in  $\mathbb{C}[[\mathbf{x}]]$  and show that the ideal projectors correspond exactly to  $D$ -invariant subspaces of  $\mathbb{C}[[\mathbf{x}]]$ . These subspaces are precisely the spaces of solution of homogeneous systems of PDEs with constant coefficients, hence the acronym.

## 1.6 Putting it all together

In this paper we will explain interrelationship between the notions mentioned above, present result and some open problems that are based on intricate interplay between these diverse fields. To keep the size of the paper reasonable I had to make a choice between rigorous proofs and accessibility of the material to a reader not terribly familiar with algebraic geometry (such as myself). I chose the latter. Thus many of the proofs are going to be only highlighted (with detailed references to the original papers). Instead I will attempt to illustrate the results with generous serving of concrete examples. **Bon appetit!**

# 2 Antipasto

## 2.1 de Boor’s equation

An obvious property of multiplication on  $\mathbb{C}[\mathbf{x}]/J$ :

$$[f[g]] = [fg] \quad (2.1)$$

translates into the following characterization of ideal projectors :

**Theorem 2.1** (*Carl de Boor [Bo1]*) *A linear operator  $P : \mathbb{C}[\mathbf{x}] \rightarrow \mathbb{C}[\mathbf{x}]$  is an ideal projector if and only if*

$$P(fg) = P(f \cdot P(g)) \quad (2.2)$$

for all  $f, g \in \mathbb{C}[\mathbf{x}]$ .

This theorem implies that unlike an arbitrary projector onto  $G$ , an ideal projector is completely determined by its values on a small set of polynomials. This makes sense. Imagine that you know that  $P$  is a Lagrange interpolation projector onto the space of polynomial of degree less than  $N$  in one variable. Knowing  $Px^N$  we can form a polynomial  $x^N - Px^N$  which has exactly  $N$  zeroes. These zeroes are the interpolation nodes (sites) for  $P$ , thus  $Px^N$  determines  $P$ . In algebraic term the ideal  $\langle x^N - Px^N \rangle$  generated by  $x^N - Px^N$  is the kernel of  $P$ .

## 2.2 The border bases

Here is the general situation:

Let  $\mathbf{g} = (g_1, \dots, g_N)$  be a linear basis for  $G$ . We define the border of  $\mathbf{g}$  to be

$$\partial\mathbf{g} := \{1, x_i g_k, \ i = 1, \dots, d, \ k = 1, \dots, N\} \setminus G. \quad (2.3)$$

For every  $J \in \mathfrak{I}_G$ , the decomposition (1.1) induces an *ideal projector*  $P_J$  onto  $G$  with  $\ker P_J = J$ . From (1.1) it follows that for every ideal  $J \in \mathfrak{I}_G$  and for every  $b \in \partial\mathbf{g}$  there exists a unique (!) polynomial  $p_b = P_J b \in G$  such that  $b - p_b \in J$ . As it turns out, the set  $\{b - p_b, b \in \partial\mathbf{g}\}$  forms an ideal basis for  $J$ , called a (generalized) border basis.

**Proposition 2.2** *Let  $J \in \mathfrak{I}_G$  and for every  $b \in \partial\mathbf{g}$  let  $p_b := P_J b$  be the unique polynomial in  $G$  such that  $b - p_b \in J$ . Then*

- (i)  $\{b - p_b, b \in \partial\mathbf{g}\}$  forms an ideal basis for  $J$ .
- (ii) If  $Pf = \sum_{\alpha \in \mathbb{Z}_+^d} a_\alpha(f) x^\alpha \in G$  then the coefficients  $a_\alpha(f)$  are polynomials

in the coefficients of polynomials  $\{Pb, b \in \partial\mathbf{g}\}$ .

The simple proof of (i) can be found in [Bo1] and equally simple proof for (ii) is in [S5].

What about a converse? That is, what polynomials  $(p_b, b \in \partial\mathbf{g}) \subset G$  have the property that the ideal  $\langle b - p_b, b \in \partial\mathbf{g} \rangle$  is in  $\mathfrak{I}_G$ ? This is the question first dealt with in [Mo] with some additional assumptions on  $G$  (cf. also [Bo3] and [KR, 6.4B], ). The extension of their results is presented below.

Mimicking the terminology of [KR, 6.4B], we will characterize those border prebases that are border bases. We will present necessary and sufficient conditions on polynomials  $\{p_b, b \in \partial\mathbf{g}\}$  for  $\{b - p_b, b \in \partial\mathbf{g}\}$  to be a basis for an ideal in  $\mathfrak{I}_G$ . As in [KR, 6.4B], the criterion involves formal multiplication operators  $M_j : G \rightarrow G$  defined by

$$M_i g_k = \begin{cases} x_i g_k & \text{if } x_i g_k \in G \\ p_{x_i g_k} & \text{if } x_i g_k \notin G \end{cases} \quad (2.4)$$

Here is the main theorem of this section:

**Theorem 2.3** *Let  $(p_b, b \in \partial\mathbf{g})$  be a sequence of polynomials in  $G$ . Then the ideal  $\langle f - p_f, f \in \partial\mathbf{g} \rangle \in \mathfrak{I}_G$  if and only if*

- (i)  $M_i M_k = M_k M_i$  for all  $i, k = 1, \dots, d$ ,
- (ii)  $g(M_1, \dots, M_d) p_1 = g$  for all  $g \in G$ .

**Proof.** First assume that  $J = \langle b - p_b, f \in \partial\mathbf{g} \rangle \in \mathfrak{I}_G$  and let  $P_J$  be the ideal projector onto  $G$  with  $\ker P_J = J$ . Then  $M_i g = P_J(x_i g)$  for all  $i = 1, \dots, d$ . It follows from (2.2) that

$$M_j M_k(g) = P_J(x_j P_J(x_k g)) = P_J(x_j x_k g) = P_J(x_k x_j g) = M_k M_j(g)$$

which proves (i). Also observe that if  $g = \sum a_\alpha \mathbf{x}^\alpha$ , then, for  $\mathbf{M} := (M_1, \dots, M_d)$  and  $g_0 := P1$ , we have

$$\begin{aligned} g(M_1, \dots, M_d)g_0 &= g = \sum a_\alpha \mathbf{M}^\alpha (P_J 1) = \sum a_\alpha P_J(x^\alpha (P_J 1)) \\ &= \sum a_\alpha P_J(x^\alpha) = P_J(\sum a_\alpha x^\alpha) = P_J g = g \end{aligned}$$

which proves (ii).

Now, suppose that (i) and (ii) holds. Then the mapping  $\varphi : \mathbb{C}[\mathbf{x}] \rightarrow \mathbb{C}[\mathbf{x}]$  defined by

$$\varphi f = f(\mathbf{M})p_1$$

is a ring homomorphism, hence its kernel

$$K := \ker \varphi = \{f \in \mathbb{C}[\mathbf{x}] : f(\mathbf{M})p_1 = 0\} \stackrel{\text{by (ii)}}{=} \{f \in \mathbb{C}[\mathbf{x}] : f(\mathbf{M}) = 0\}$$

is an ideal in  $\mathbb{C}[\mathbf{x}]$ . By (ii) the range of  $\varphi$  is  $G$  and  $K \cap G = 0$ . By the fundamental theorem of homomorphisms  $\mathbb{C}[\mathbf{x}]/K$  is isomorphic to  $G$ . In particular codimension of  $K$  is equal to  $\dim G$  and  $K$  complements  $G$ .

Let  $h_b$  be the unique element in  $G$  such that  $b - h_b \in K$ . We need to show that  $J = K$  or, alternatively that  $h_b = g_b$  for every  $b \in \partial \mathfrak{g}$ . Since  $b - h_b \in K$  we have

$$0 = (b(\mathbf{M}) - h_b(\mathbf{M}))p_1 \stackrel{\text{by (ii)}}{=} b(\mathbf{M})p_1 - h_b$$

On the other hand, by definition of  $\mathbf{M}$ , we have  $b(\mathbf{M})p_1 = p_b$  which implies that  $p_b = h_b$  for all  $b \in \partial \mathfrak{g}$ . ■

**Remark 2.4** *If  $G$  is a  $D$ -invariant subspace of  $\mathbb{K}[\mathbf{x}]$  spanned by monomials, then  $1 \in G$  and, by the  $D$ -invariance, the condition (ii) of the Theorem 2.3 is automatically satisfied (see example section 3). Hence the theorem 2.3 generalizes theorem 6.4.30 of [KR] with, what seems to be, a shorter, simpler proof, courtesy of the language of ideal projectors.*

### 2.3 The border scheme

The operators  $M_1, \dots, M_k$  can be written as  $N \times N$  matrices in the basis  $\mathfrak{g}$  and, the polynomial  $p_1 \in G$  generates an  $N \times 1$  matrix of its coefficients.

**Definition 2.5** *The affine scheme  $\mathcal{B}_{\mathfrak{g}}$  defined by the ideal  $I_{\mathfrak{g}}$  generated by the entries of the matrices  $M_j M_i - M_i M_j, i, j = 1, \dots, d$  and the coordinates of the vector  $p_1$ :*

$$\langle g_k(M_1, \dots, M_d)p_1 - g_k \rangle, k = 1, \dots, N \quad (2.5)$$

*is called the generalized border scheme for  $\mathfrak{g}$  or  $\mathfrak{g}$ -border scheme. It parametrizes the family of ideals  $\mathfrak{I}_G$  or, equivalently, the family of ideal projectors  $\mathfrak{P}_G$ .*

**Proposition 2.6** *It is clear from construction ideal projector  $P \in \mathfrak{P}_G$  (ideals  $J \in \mathfrak{I}_G$ ) are in one-to-one correspondence with the points in  $\mathcal{B}_{\mathfrak{g}}$ .*

Thus we will sometimes refer to  $P \in \mathfrak{P}_G$  ( $J \in \mathfrak{I}_G$ ) as a point  $P \in \mathfrak{P}_G$ .

### 3 Aperitif

In this section we will illustrate how the notions of the previous sections apply for concrete spaces  $G$ . About the easiest multivariate space one can find, is the space of linear function in  $\mathbb{C}[x, y]$ . As you will see, even in this case the computations are quite involved. Nevertheless, they are worth going through. Once accustomed to it, this example acts like a Led Zeppelin record: when played backwards, it sends you messages about the general theory. (For the univariate theory see [S1]).

We will now attempt to determine all ideal projectors onto the three dimensional subspace  $G \subset \mathbb{C}[x, y]$  spanned by its basis  $\mathbf{g} = (1, x, y)$ .

By the theorem 2.3, to describe ideal projectors  $P$  onto  $G$  we only need know the values of  $P$  on the set of monomials  $x^2, xy$  and  $y^2$ . In other words the polynomials  $x^2 - Px^2$ ,  $xy - P(xy)$  and  $y^2 - Py^2$  will form an ideal basis for the ideal  $\ker P$ .

Assume that  $P$  is an ideal projector onto  $G$  and

$$\begin{aligned} Px^2 &= a_0 + b_0x + c_0y, \\ Pxy &= a_1 + b_1x + c_1y, \\ Py^2 &= a_2 + b_2x + c_2y. \end{aligned} \tag{3.1}$$

We need to find conditions on nine coefficients  $(a_0, a_1, a_2, b_0, b_1, b_2, c_0, c_1, c_2)$  that guarantee that the ideal

$$\langle x^2 - Px^2, xy - Pxy, y^2 - Py^2 \rangle$$

complements  $G$ ?

To answer this question we form formal multiplication matrices

$$M_1 = \begin{bmatrix} 0 & a_0 & a_1 \\ 1 & b_0 & b_1 \\ 0 & c_0 & c_1 \end{bmatrix}; M_2 = \begin{bmatrix} 0 & a_1 & a_2 \\ 1 & b_1 & b_2 \\ 0 & c_1 & c_2 \end{bmatrix} \tag{3.2}$$

and use Theorem 2.3. For our choice of  $G$ , the conditions (ii) of the theorem is automatically satisfied. All that is left is to enforce the commutativity. The six quadratic equations obtained from  $M_1M_2 - M_2M_1 = 0$  are

$$\left\{ \begin{array}{l} (a_0b_1 + a_1c_1) - (a_1b_0 + a_2c_0) = 0, \\ (a_1 + b_0b_1 + b_1c_1) - (b_0b_1 + b_2c_0) = 0, \\ (c_1^2 + b_1c_0) - (a_0 + b_0c_1 + c_0c_2) = 0, \\ (a_0b_2 + a_1c_2) - (a_1b_1 + a_2c_1) = 0, \\ (a_2 + b_0b_2 + b_1c_2) - (b_1^2 + b_2c_1) = 0, \\ (b_2c_0 + c_1c_2) - (a_1 + b_1c_1 + c_1c_2) = 0. \end{array} \right.$$

A close examination reveals that there are a lot of redundancy among these equation. The solutions to these equations are given by

$$\begin{aligned} a_0 &= -b_0c_1 + c_1^2 + b_1c_0 - c_0c_2, \\ a_1 &= b_2c_0 - b_1c_1, \\ a_2 &= b_1^2 - c_2b_1 - b_0b_2 + b_2c_1. \end{aligned} \tag{3.3}$$

The border scheme  $\mathcal{B}_9$  is a six-dimensional variety affine variety in  $\mathbb{C}^9$  that consists of all nine-tuples

$$(a_0, a_1, a_2, b_0, b_1, b_2, c_0, c_1, c_2) \quad (3.4)$$

satisfying (3.3).

By checking (3.3) we see that the following four projectors define by

$$\begin{aligned} T : Tx^2 = T(xy) = Ty^2 = 0, \\ P_* : P_*x^2 = y, P_*(xy) = P_*y^2 = 0, \\ L : Lx^2 = x, L(xy) = 0, Ly^2 = y, \\ H : Hx^2 = Hxy = Hy^2 = y \end{aligned} \quad (3.5)$$

are in fact ideal projectors onto  $G$ . The first,  $T$  is the Taylor projector onto  $G$ , it interpolates  $f \in \mathbb{k}[x, y]$  and its first partial derivatives at 0. The dual space  $(\ker P_*)^\perp$  is a  $D$ -invariant subspace of  $\mathbb{k}[[x, y]]$  spanned by  $1, x, x^2 + 2y$ . Thus  $P_*$  also interpolates at zero various derivatives at zero, namely

$$\delta_0, \delta_0 \circ D_x, \delta_0 \circ (D_x^2 + 2D_y)$$

and is a different projector. Hence, unlike the case in one variable there are two (infinitely many) ideal projectors onto  $G$  such that  $\mathcal{Z}(\ker P) = \{0\}$ . The projector  $L$  is a Lagrange projector interpolating at sites  $(0, 0), (1, 0)$  and  $(0, 1)$ . The dual space  $(\ker L)^\perp$  is a span of the power series expansion of three distinct exponential functions:  $1 = e^0, e^x$  and  $e^y$ . Finally the last projector  $H$  interpolates at zero, at  $(1, 1)$  and the derivative  $D_x$  at zero.  $(\ker H)^\perp = \text{span}\{e^0, xe^0, e^{x+y}\}$ .

## 4 Soup and salad

### 4.1 Cyclic commuting matrices

Given  $P \in \mathfrak{P}_G$  and  $i = 1 : d$ , we define multiplication operators  $M_i : G \rightarrow G$  by

$$M_i(g) = P(x_i g). \quad (4.1)$$

These operators are similar (literally and figuratively) to the multiplication maps  $m_j$  on  $\mathbb{k}[\mathbf{x}]/J$  defined by  $m_j([f]) := [x_j f] \in \mathbb{k}[\mathbf{x}]/J$  for every  $[f] \in \mathbb{k}[\mathbf{x}]/J$ . A relationship between ideals, multiplication maps and numerical analysis was initiated and explored by H. Stetter [St]. Observe that these operators are precisely the operators defined in (2.4) with  $p_{x_i g_k} = P(x_i g_k)$ . Therefore the sequence  $\mathbf{M}_P := (M_1, \dots, M_d)$  is a cyclic commuting sequence with the cyclic vector  $g_0 := P(1)$ .

Nearly all the information about  $P$  can be read off the sequence  $\mathbf{M}_P$ :

**Proposition 4.1** *Let  $\varphi : \mathbb{C}[\mathbf{x}] \rightarrow \mathbb{C}[\mathbf{x}]$  be defined by  $\varphi(f) = f(\mathbf{M}_P)g_0 \in \mathbb{C}[\mathbf{x}]$ . Then*

- (i)  $\ker P = \ker \varphi = \{f \in \mathbb{C}[\mathbf{x}] : f(\mathbf{M}_P)g_0 = 0\} = \{f \in \mathbb{C}[\mathbf{x}] : f(\mathbf{M}_P) = 0\}$ .
- (ii) The restriction  $\varphi|_G$  of  $\varphi$  to  $G$  is an isomorphism on  $G$ .
- (iii)  $P = \left(\varphi|_G\right)^{-1} \circ \varphi$

The converse to this is also true (cf. [BS]):

**Theorem 4.2** *Let  $\mathbf{L} := (L_1, \dots, L_d)$  be a cyclic sequence of commuting  $N \times N$  matrices with a cyclic vector  $v_0 \in \mathbb{C}^N$ . Let  $\varphi : \mathbb{C}[\mathbf{x}] \rightarrow \mathbb{C}[\mathbf{x}]$  be defined by  $\varphi(f) = f(\mathbf{L})v_0 \in \mathbb{C}^N$ , let*

$$J_{\mathbf{L}} := \{f \in \mathbb{C}[\mathbf{x}] : f(\mathbf{M}_P) = 0\} \quad (4.2)$$

*Then  $\mathbf{L}$  is similar to the matrices of multiplication operators  $\mathbf{M}_P$  of any ideal projector  $P$  with  $\ker P = J_{\mathbf{L}}$ .*

## 4.2 Duality and PDEs

The space  $\mathbb{C}[[\mathbf{x}]] \supset \mathbb{C}[\mathbf{x}]$  is the space of all formal power series in  $d$  variables. A generic element  $f \in \mathbb{C}[[\mathbf{x}]]$  is written as a formal sum

$$f(\mathbf{x}) = \sum \hat{f}(\boldsymbol{\alpha}) \mathbf{x}^{\boldsymbol{\alpha}} \quad (4.3)$$

where  $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$  runs through all multiindices in  $\mathbb{Z}_+^d$  and  $\mathbf{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \dots x_d^{\alpha_d}$ . Given  $f \in \mathbb{C}[[\mathbf{x}]]$  and a sequence  $\mathbf{L} = (L_1, \dots, L_d)$  of commuting operators on some linear space  $V$ , we define a formal operator  $f(\mathbf{L})$  on  $V$  to be

$$f(\mathbf{L})v = \sum \hat{f}(\boldsymbol{\alpha}) L_1^{\alpha_1} \dots L_d^{\alpha_d} v.$$

In particular if  $D_i$  are operators of differentiation on  $V = \mathbb{C}[\mathbf{x}]$  with respect to the variable  $x_i$  and  $\mathbf{D} := (D_1, \dots, D_d)$  then for every  $F \in \mathbb{C}[[\mathbf{x}]]$  and every  $f \in \mathbb{C}[\mathbf{x}]$  the pairing

$$\langle F, f \rangle := (\bar{F}(\mathbf{D})f)(0) = \sum \boldsymbol{\alpha}! \bar{F}(\boldsymbol{\alpha}) \hat{f}(\boldsymbol{\alpha})$$

defines an isomorphism between  $\mathbb{C}[[\mathbf{x}]]$  and the algebraic dual  $\mathbb{C}'[\mathbf{x}]$  of  $\mathbb{C}[\mathbf{x}]$ . It is easy to see that

$$\langle F, x_i f \rangle = \langle D_i F, f \rangle, \quad (4.4)$$

in other words  $D_i$  on  $\mathbb{C}[[\mathbf{x}]]$  is an adjoint of the operator of multiplication by  $x_i$  on  $\mathbb{C}[\mathbf{x}]$ .

**Example:** It is easy (and insightful) to check that the point evaluation functional  $\delta_{\mathbf{z}}: \delta_{\mathbf{z}}(f) = f(\mathbf{z})$  on  $\mathbb{C}[\mathbf{x}]$  corresponds to the functional  $F \in \mathbb{C}[[\mathbf{x}]]$  defined by the expansion of  $e^{\mathbf{z} \cdot \mathbf{x}}$  in powers of  $\mathbf{x}$ .

The following theorem due to Macauley [Ma] follows immediately from (4.4):

**Theorem 4.3** *A subspace  $J \subset \mathbb{C}[\mathbf{x}]$  is an ideal iff  $J^\perp \subset \mathbb{C}[[\mathbf{x}]]$  is  $D$ -invariant, i.e.,  $F \in J^\perp$  implies  $D_i F \in J^\perp$ .*

Thus the study of  $\mathfrak{J}_G$  (hence  $\mathfrak{J}_G$ ) is equivalent to the study of  $D$ -invariant subspaces  $\Phi \subset \mathbb{C}[[\mathbf{x}]]$  that are correct for  $G$ .

**Example:** Keeping in mind the example above, we conclude that the Lagrange interpolation projectors are determined by subspaces of  $\mathbb{C}[[\mathbf{x}]]$  which are



spanned by pure exponentials:  $\Phi = \text{span}\{e^{\mathbf{z}^j \cdot \mathbf{x}} : j = 1 : N\}$  while the Taylor projector is determined by a span of pure polynomials:  $\Phi = \text{span}\{\mathbf{x}^\alpha : |\alpha| \leq n\} =: \mathbb{C}[\mathbf{x}]_{\leq n}$ .

As the multiplication matrices are associated with the ideal projector, the adjoints of those matrices determine the rules of differentiating functions in  $(\ker P)^\perp$ . Let  $J \in \mathfrak{I}_G$ . For every  $F \in J^\perp$  we denote the restriction of  $F$  to  $G$  by  $F^*$ . Since  $G$  complements  $J$ , it follows that  $J^\perp$  is an  $N$ -dimensional subspace over  $G$ , hence the dual space  $G^*$  is

$$(J^\perp)^* := \{F^*, F \in J^\perp\}. \quad (4.5)$$

If  $P_J$  is the ideal projector onto  $G$  with  $\ker P = J$ , then

$$P_J = \sum g_j \otimes F_j \quad (4.6)$$

where  $\mathbf{g} = (g_1, \dots, g_N)$  is a basis for  $G$ , and  $(F_1, \dots, F_N)$  is the dual basis in  $J^\perp$ , i.e.,  $F_j(g_k) = \delta_{j,k}$ . The adjoint operator  $P^* : \mathbb{C}[[\mathbf{x}]] \rightarrow \mathbb{C}[[\mathbf{x}]]$  is a projector  $P^* = \sum F_j \otimes g_j$  having  $J^\perp$  as its range.

**Theorem 4.4** *Let  $J \in \mathfrak{I}_G$  and let  $\mathbf{M}_J = (M_1, \dots, M_d)$  be the matrices of multiplication operators:*

$$M_i(g) = P_J(x_i g)$$

*in the basis  $\mathbf{g} = (g_1, \dots, g_N)$ . Let  $(F_1, \dots, F_N)$  be the dual basis in  $J^\perp$ , i.e.,  $F_j(g_k) = \delta_{j,k}$ . Then*

$$D_i \begin{bmatrix} F_1 \\ \vdots \\ F_N \end{bmatrix} = M_i^* \begin{bmatrix} F_1 \\ \vdots \\ F_N \end{bmatrix} \quad (4.7)$$

**Proof.** For every  $g \in G$  we have  $F(M_i g) = F^*(M_i g) = (M_i^t F^*)(g)$ . On the other hand

$$F(M_i g) = F(P(x_i g)) = (P^* F)(x_i g) = F(x_i g) = (D_i F)(g) = (D_i F)^*(g),$$

where the third equality follows  $F \in J^\perp = \text{ran } P^*$ . Hence  $(M_i^t F^*) = (D_i F)^*$ .

This means that  $(D_i F_k)^* = \sum m_{j,k}^{(i)} F_j^*$ , where  $m_{j,k}^{(i)}$  is the  $j, k$ -entry in the matrix  $M_i^*$ . On the other hand, by the  $\partial$ -invariance,  $D_i F_k = \sum a_{j,k}^{(i)} F_j$  for some coefficients  $a_{j,k}^{(i)}$ . Since  $(F_j^*)$  is a basis in  $G^*$  it follows that  $a_{j,k}^{(i)} = m_{j,k}^{(i)}$ . ■

### 4.3 Primary ideals

An ideal  $J \subset \mathbb{C}[\mathbf{x}]$  is called primary if  $fg \in J$  implies  $f^m \in J$  for some  $m \in \mathbb{N}$  or  $g \in J$ .

**Theorem 4.5** *Let  $J$  be a zero-dimensional ideal in  $\mathbb{C}[\mathbf{x}]$ . Then the following are equivalent:*

(CA)  *$J$  is primary*

- (G)  $\mathcal{Z}(J) = \{\mathbf{z}\}$ , i.e.,  $\mathcal{Z}(J)$  consists of one point  
(PDE)  $J^\perp = e^{\mathbf{z} \cdot \mathbf{x}} M$  where  $M \subset \mathbb{C}[\mathbf{x}]$  is a  $d$ -invariant subspace of polynomials  
(!).  
(LA)  $\sigma(\mathbf{M}_L) = \{\mathbf{z}\}$ .

The equivalence of (CA) and (G) is standard (cf. [CLO1], [BR]). The (PDE) was explored in [M] (cf. also [Bo1]). The (LA) follows from [MSh].

#### 4.4 Primary decomposition

**Definition 4.6** Let  $\mathbf{L} := (L_1, \dots, L_d)$  be a  $d$ -tuple of operators on  $V$ . A direct sum decomposition

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_t \quad (4.8)$$

is  $\mathbf{L}$ -invariant if each subspace  $V_k$ ,  $k = 1, \dots, t$  is an invariant subspace for each of the operators  $L_j$ ,  $j = 1, \dots, d$ .

Letting  $L_{j,k} := L_j|_{V_k}$  denote the restriction of  $L_j$  onto  $V_k$  we write

$$\mathbf{L}_k = \mathbf{L}|_{V_k} := (L_{1,k}, \dots, L_{d,k}). \quad (4.9)$$

The simultaneous block-diagonalization of  $\mathbf{L}$  into  $t$  blocks amounts to nothing more than the  $\mathbf{L}$ -invariant decomposition (4.8) of  $V$ : Indeed, for an appropriately chosen bases, the matrix  $\tilde{L}_j$  of  $L_j$  can be written in a block-diagonal form

$$\tilde{L}_j = \begin{bmatrix} \tilde{L}_{j,1} & 0 & \cdots & 0 \\ 0 & \tilde{L}_{j,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{L}_{j,t} \end{bmatrix}, \quad (4.10)$$

A  $d$ -tuple  $\boldsymbol{\lambda} := (\lambda_1 \dots \lambda_d) \in \mathbb{C}^d$  is called an eigentuple of  $\mathbf{L}$  if there exists a non-zero vector  $v \in V$  such that  $L_j v = \lambda_j v$  for all  $j = 1 : d$ . The set of all eigentuples of  $\mathbf{L}$  is called the joint spectrum of  $\mathbf{L}$  and is denoted by  $\sigma(\mathbf{L})$ . The next proposition seems to be well-known among experts. For a cute proof we refer to [MSh]:

**Proposition 4.7** Let  $\mathbf{L}$  be a  $d$ -tuple of pairwise commuting operators on  $V$ . Then

- (i)  $\sigma(\mathbf{L}) = \mathcal{Z}(J_{\mathbf{L}})$ .  
(ii) There exists an  $\mathbf{L}$ -invariant decomposition of  $V$ :

$$V = V_1 \oplus V_2 \oplus \dots \oplus V_{\#\sigma(\mathbf{L})}. \quad (4.11)$$

Adding the assumption that  $\mathbf{L}$  is cyclic gives us more:

**Theorem 4.8** (cf. [MSh]) If  $\mathbf{L}$  is cyclic then (4.11) is a unique maximal  $\mathbf{L}$ -invariant decomposition of  $V$ :

$$V = \bigoplus_{\boldsymbol{\lambda} \in \sigma(\mathbf{L})} V_{\boldsymbol{\lambda}},$$

i.e., the space  $V$  cannot be decomposed into more than  $\#\sigma(\mathbf{L})$  of  $\mathbf{L}$ -invariant subspaces. More over  $\sigma(\mathbf{L}|_{V_\lambda}) = \{\lambda\}$ , hence consists of a single eigentuple.

The reason is that if  $\mathbf{L}$  is cyclic then eigenvectors for  $\mathbf{L}$  in different  $V_k$  and  $V_j$  correspond to different eigentuples.

**Theorem 4.9** *Let  $J$  be a zero-dimensional ideal with  $\mathcal{Z}(J) = \{\mathbf{z}_1, \dots, \mathbf{z}_m\}$ . Then*

*(LA) There exists unique (up to the order) maximal  $\mathbf{M}_J$ -invariant decomposition*

$$G = G_1 \oplus G_2 \oplus \dots \oplus G_m \quad (4.12)$$

*and  $\sigma(\mathbf{M}_{J|G_j}) = \{\mathbf{z}_j\}$ .*

*(AT) The ideal projector  $P_J$  has a unique maximal decomposition as a sum of  $m$  ideal projectors:*

$$P = P_1 + \dots + P_m \quad (4.13)$$

*where each  $P_j$  is an ideal projector onto  $G_j$  interpolating at exactly one point and  $P_k P_j = \delta_{k,j} P_j$ .*

*(CA: Lasker-Noether) There is unique (minimal with respect to containment) primary decomposition*

$$J = \cap J_j \quad (4.14)$$

*with each  $J_j$  is a primary ideal with  $\mathcal{Z}(J_j) = \{\mathbf{z}_j\}$ .*

*(PDE) The subspace  $J^\perp \subset \mathbb{C}[[\mathbf{x}]]$  has a unique maximal decomposition*

$$J^\perp = e^{\mathbf{z}_1 \cdot \mathbf{x}} \cdot H_1 \oplus \dots \oplus e^{\mathbf{z}_m \cdot \mathbf{x}} \cdot H_m \quad (4.15)$$

*where each  $H_j$  is a  $D$ -invariant subspace of polynomials.*

The (CA) is the famous Lasker-Noether theorem, (PDE) was observed by [M] (cf. also [Bo1], [BR]) and follows from (CA) since (4.14) implies  $J^\perp = \oplus J_j^\perp$ . (LA) and (AT) are from [MSh].

## 5 Appetizers

### 5.1 Radical ideals

An ideal  $J \subset \mathbb{C}[x]$  is called a radical ideal if  $f^m \in J$  implies  $f \in J$ .

**Theorem 5.1** *Let  $J \subset \mathbb{C}[x]$  be a zero-dimensional ideal. Then the following are equivalent:*

*(A):  $J$  is a radical ideal.*

*(G):  $\#\mathcal{Z}(J) = \text{codim} J$*

*(PDE):  $J^\perp$  is a linear span of pure exponentials:  $J^\perp = \text{span}\{e^{\mathbf{z}_k \cdot \mathbf{x}}, k = 1 : N\}$ .*

*(LA): The matrices  $(M_1 \dots M_d) = \mathbf{M}_J$  are simultaneously diagonalizable. Moreover, the diagonal elements consist of interpolation sites for the ideal projector  $P_J$ .*

*(AT):  $P_J$  is a Lagrange projector interpolating at sites  $\{\mathbf{z}_1 \dots \mathbf{z}_N\}$*

**Proof.** Again, equivalence of (A) and (G) is standard (cf. [CLO1]). (PDE) follows from Theorem 4.5. (LA) was first observed in [St], cf. also [Bo1] and [BS]; but now follows from Theorem 4.8. ■

**Example:** The projector  $L$  from (3.5) is a Lagrange projector interpolating at sites  $(0, 0)$ ,  $(1, 0)$  and  $(0, 1)$ . The multiplication matrices for this projector are

$$\mathbf{M}_L = \left( \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 1 \end{bmatrix} \right) \sim \left( \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right)$$

that is the matrices of  $\mathbf{M}_L$  are diagonalizable.  $J^\perp = \text{span}\{1 = e^0, e^x, e^y\}$  which is the span of pure exponential.

**Remark 5.2** *The equivalence of (A) and (LA) of the theorem was first observed by Hans Stetter [St]. He also noticed that the joint spectrum (diagonal pairs) of  $\mathbf{M}_L$  are precisely the interpolation sites, which also follows from the Theorem 4.8. More over the eigenvectors corresponding to these eigentuple are basic Lagrange polynomials that vanish on all points but one.*

## 5.2 Curvilinear ideals

An ideal  $J$  is called curvilinear if there exists a linear form  $X = \sum_{j=1}^d \alpha_j x_j$  such that  $J$  complements the space  $\text{span}\{1, X, \dots, X^{N-1}\}$ .

**Theorem 5.3** *For an ideal  $J \in \mathfrak{J}_G$  the following are equivalent:*

- (AT):  $J$  is curvilinear
- (LA): The matrix  $M := \sum_{j=1}^d \alpha_j M_j$  is non-derogatory.
- (CA) Every ideal  $J_k$  in the primary decomposition  $J = \cap_{k=1}^s J_k$  is curvilinear.
- (PDE): If  $J^\perp = \oplus_{k=1}^{\#Z(J)} (e^{\mathbf{z}_k \cdot \mathbf{x}} H_k)$  then each  $H_k$  contains at most one linear polynomial.

**Proof.** Let  $G_X := \text{span}\{1, X, \dots, X^{N-1}\}$  and let  $Q$  be an ideal projector onto  $G_X$  with  $\mathbf{M}_Q = (L_1, \dots, L_d)$  and define  $L_X := \sum a_i L_i$ . By the Theorem 2.3,  $(1, L_X 1, \dots, L_X^{N-1} 1)$  spans the space  $G_X$ , hence  $\sum a_i L_i$  is non-derogatory. By the theorem 2.7,  $\mathbf{M}_Q$  is similar to  $\mathbf{M}_P$  hence  $\sum a_i M_i$  is similar to  $L_X$  hence non-derogatory. This proves (LA). The converse: (LA) $\Rightarrow$ (AT) immediately follows from Proposition 4.1. To prove (CA), let  $\mathbf{M}_J = \text{diag}(\mathbf{M}_J^{(k)})$  and  $M_i = \text{diag}(M_{i,k})$  as in (4.10). Then  $\sum a_i M_i = \sum a_i \text{diag}(M_{i,k})$  is non-derogatory if and only if  $\sum_i a_i M_{i,k}$  is non-derogatory for each  $k$ .

Finally, to prove the equivalence of (PDE) to the rest of the statements of the theorem, observe that, by (CA) it is enough to prove this for primary ideals  $J$  with  $\mathcal{V}(J) = \{0\}$ , i.e., with  $J^\perp \subset \mathbb{C}[\mathbf{x}]$ . Without loss of generality, assume

that  $J \in \mathfrak{J}_G$  with  $G = \text{span}\{1, x_1, \dots, x_1^{N-1}\}$ . This means that the dual basis for  $J^\perp$  is of the form

$$\{x_1^k + f_k, k = 0, \dots, N-1\} \quad (5.1)$$

where  $f_k$  do not contain the monomials  $\{1, x_1, \dots, x_1^{N-1}\}$ . In particular, it (5.1) contains at most one linear form:  $x_1 + f_1$ .

Conversely, assume that  $x_1$  is the only linear form in  $J^\perp$  and  $J$  is not curvilinear. Then  $J^\perp$  could not have a basis of the form (??) hence no polynomial in  $J^\perp$  can contain a monomial  $x_1^m$  for  $m \geq N-1$ , for otherwise its consecutive derivatives would produce a basis of the form (3.5). Since  $J^\perp$  is  $N$ -dimensional, it must contain a polynomial  $F$  that has no pure powers of  $x_1$  in it. But that means that an appropriate derivative of  $F$  will give a linear form  $x_k$  with  $k \neq 1$ . Indeed let  $u = x_1^{\alpha_1} \dots x_k^{\alpha_k} \dots x_d^{\alpha_d}$  be a monomial of the highest degree in  $F$  such that  $\alpha_k \geq 1$  for some  $k > 1$ . Then

$$x_k = (D_1^{\alpha_1} \dots D_k^{\alpha_k-1} \dots x_d^{\alpha_d})u = (D_1^{\alpha_1} \dots D_k^{\alpha_k-1} \dots x_d^{\alpha_d})F \in J^\perp$$

which gives the desired contradiction. ■

**Corollary 5.4** *Every radical ideal projector is curvilinear. Converse is not true.*

**Proof.** Diagonalizing  $\mathbf{M}_J$  and taking a generic linear combination  $\sum a_i M_i$  we conclude that  $S(\sum a_i M_i)S^{-1}$  has distinct elements on the diagonal, hence is non-derogatory. Therefore  $\sum a_i M_i$  is non-derogatory and  $J$  is curvilinear. The second part of the statement follows from the example below. ■

**Example:** The kernels of projectors  $P_*$ ,  $L$  and  $H$  are curvilinear ideals. In particular, looking at the duals we see that  $(\ker P_*)^\perp$  contains precisely one linear term while none of the  $H_j$  for  $(\ker L)^\perp$  contain linear terms since  $(\ker L)^\perp$  is a combination of pure exponentials. The ideal  $(\ker P_*)^\perp$  is not radical, yet curvilinear. It complements the  $\text{span}\{1, y, y^2\}$ . The  $(\ker H)^\perp$  has one summand with the linear term and one without. Both ideals  $(\ker H)^\perp$  and  $(\ker L)^\perp$  complement the space, say,  $\{1, x + y, (x + y)^2\}$ . The kernel of the Taylor projector  $\ker T$  is not curvilinear since  $(\ker T)^\perp$  contains two linear forms.

**Remark 5.5** *The equivalence of (AT) and (LA) was observed in [BS] and [Co]. I learned the equivalence of (AT) and (PDE) from [Co]. The proof presented here is quite different, and in my opinion, much simpler than the one in [Co].*

### 5.3 Gorenstein Ideals

**Definition 5.6** *A zero-dimensional ideal  $J \in \mathfrak{J}_G$  is called Gorenstein if there exists one function  $F \in \mathbb{C}[[x]]$  such that  $\mathcal{D}(F) = J^\perp$ .*

Gorenstein ideals form an important class of ideal in commutative algebra (cf. [B]).

**Theorem 5.7** *Let  $J \in \mathfrak{J}_G$  be a zero-dimensional ideal. The following are equivalent:*

- (PDE)  $J$  is Gorenstein.
- (LA1) The sequence of adjoints  $\mathbf{M}_J^* = (M_1^*, \dots, M_d^*)$  is cyclic.
- (LA2)  $\mathbf{M}_J^*$  is similar to  $\mathbf{M}_J$
- (CA) Every ideal  $J_k$  in the primary decomposition  $J = \cap_{k=1}^s J_k$  is Gorenstein.

**Proof.**  $\mathbf{M}^*$  is cyclic if and only if there exists an  $F \in J^\perp$  such that  $\{p(\mathbf{M}^*)F^*, p \in \mathbb{C}[\mathbf{x}]\} = G^*$ . This is equivalent to  $\{(p(\mathbf{D})F)^*, p \in \mathbb{C}[\mathbf{x}]\} = G^*$  which, in turn, is equivalent to  $\mathcal{D}(F) = \{(p(\mathbf{D})F), p \in \mathbb{C}[\mathbf{x}]\} = J^\perp$ .

To prove (LA2) observe the ideals generated by  $\mathbf{M}_J^*$  and  $\mathbf{M}_J$  by (4.2) are equal. Thus, by the theorem 4.2 the sequence  $\mathbf{M}_J^*$  is similar to  $\mathbf{M}_J$ . ■

**Example 5.8** *The multiplication operators of the Taylor projector  $T$  from (3.5) are*

$$\mathbf{M}_T = \left( \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right), \mathbf{M}_T^* = \left( \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)$$

and  $\mathbf{M}_T^*$  is not a cyclic sequence, as was noted in [BS]. The reason for this is that  $(\ker T) = \text{span}\{1, x, y\}$  is not a deflation of a single polynomial. On the other hand the projector  $P_*$ , being curvilinear is Gorenstein

$$\mathbf{M}_{P_*} = \left( \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right), \mathbf{M}_{P_*}^* = \left( \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \right)$$

and  $\mathbf{M}_{P_*}^*$  is a cyclic sequence with cyclic vector  $(0, 0, 1)$ . Observe that the dual  $(\ker P_*)^\perp$  are given by  $(1, x, 2y + x^2)$  and  $\mathcal{D}(2y + x^2) = (\ker P)^\perp$ .

Finally consider  $F = xy$ , then  $\mathcal{D}(F) = \text{span}\{1, x, y, xy\}$  and

$$J = \{f \in \mathbb{C}[x, y] : P(f) = 0, P \in \mathcal{D}(F)\} \quad (5.2)$$

is Gorenstein, yet not curvilinear, since  $J^\perp$  contains two linearly independent polynomials:  $x$  and  $y$  of degree 1.

**Proposition 5.9** *If  $J$  is curvilinear, then  $J$  is Gorenstein.*

**Proof.** If  $J$  is curvilinear then there exists a linear combination  $M_0 := \sum a_j M_j$  such that  $M_0$  is non-derogatory. This implies that  $\mathbf{M}$  is non-derogatory, hence cyclic. Thus  $\mathbf{M}^* = (M_j^*)$  is cyclic and  $J$  is Gorenstein. ■

Theorem 5.7 has an interesting and unexpected corollary to linear algebra:

Every square matrix  $M$  is similar to its adjoint. This is not the case for sequences of commuting matrices as is seen in the example where  $\mathbf{M}_T^*$  is not cyclic. From equivalence of (LA1) and (LA2) of the theorem 5.7 we immediately obtain the following corollary:

**Theorem 5.10** *A cyclic sequence  $\mathbf{L} = (L_1, \dots, L_d)$  is similar to its adjoint  $\mathbf{L}^* = (L_1^*, \dots, L_d^*)$  if and only if  $\mathbf{L}^*$  is cyclic.*

## 5.4 Hermite projectors

**Definition 5.11** A projector  $P \in \mathfrak{P}_G$  is called *Hermite* if there exists a sequence of Lagrange projectors  $P_n$  such that

$$P_n f \rightarrow P f \quad (5.3)$$

for every  $f \in \mathbb{C}[\mathbf{x}]$ . An ideal  $J = \ker P$  for Hermite projector  $P$  is called a *Hermite ideal*.

**Proposition 5.12** Let  $J = \cap J_k$  be a primary decomposition of the ideal  $J \in \mathfrak{J}_G$ . Then  $J$  is Hermite if and only if each  $J_k$  is.

I have a messy proof for this yet I feel that there should exist a short concise one that, so far, I failed to find.

Every Lagrange projector is Hermite. In one variable every ideal projector  $P$  is Hermite since the polynomial  $h \in \mathbb{C}[x]$  that generates the ideal  $\ker P$  can be approximated by polynomials  $h_n$  of the same degree with distinct zeroes. The ideal projectors with ideals generated by  $h_n$  are Lagrange.

## 6 Main course

My personal journey through the landscape of ideal projectors began with a question of Carl de Boer [Bo1]: Is every finite-dimensional ideal projector Hermite? Ironically there were questions in other areas of mathematics that at the end turned out to be the same or almost the same questions. Here is the list:

**Problem 6.1** (AT, [Bo1]) Is every finite-dimensional ideal projector Hermite?  
 (AG, [Fo]) Is the border schemes  $\mathcal{B}_{\mathfrak{g}}$  irreducible  
 (LA, [Ge], [Gu]) Can every  $d$ -tuple of commuting matrices be approximated by commuting diagonalizable  $d$ -tuple of matrices  
 (PDE, [Le]) Can a space of solutions of the homogeneous systems of PDEs with constant coefficients be approximated by spaces of pure exponentials.

With the hindsight of the previous chapters, you can guess that the answer to all these questions is the same. It turns out to be: “Yes” in one and two variables and “No” in three or more variables. For (AT) and (PDE) this was done in [S2], for (AG) in [Fo] and [Ia]

John Fogarty [Fo] proved a remarkable theorem, which translated from Hilbert schemes to border schemes says:

**Theorem 6.2** If  $\mathfrak{g}$  spans an  $N$ -dimensional subspace of  $\mathbb{C}[x, y]$  then  $\mathcal{B}_{\mathfrak{g}}$  is smooth, connected, irreducible affine variety of dimension  $2N$ .

Before we explain the words smooth and irreducible, let us lay the groundwork for asking questions. Are any of the statements in the theorem 6.2 true

for border schemes in more than two variables? Do the answers depend on particular space  $G$ ?

Surprisingly the answer to the first question is “No” and to the second is “Yes”. Somehow, in little understood way, the structure of the border scheme depends strongly on the subspace  $G$ .

## 6.1 Topology

Next proposition says that various topologies on  $\mathcal{B}_g$  are all the same:

**Theorem 6.3** *Let  $P_n$  and  $P$  be in  $\mathfrak{P}_G$ . The following are equivalent*

- (AT)  $P_n f \rightarrow P f$  for every  $f$
- (AG)  $P_n \rightarrow P$  as points in  $\mathcal{B}_g$
- (LA) The matrices  $\mathbf{M}_{P_n} \rightarrow \mathbf{M}_P$
- (PDE) For every  $F \in (\ker P)^\perp$  there exist  $F_n \in (\ker P_n)^\perp$  such that  $F_n(f) \rightarrow F(f)$  for every  $f \in \mathbb{C}[\mathbf{x}]$ .

**Proof.** The equivalence of (AT), (AG) and (LA) are immediate from Proposition 2.2 (ii) and constructions of  $\mathbf{M}_P$ . The equivalence of (PDE) is intuitively obvious. Since  $P_n$  and  $P$  are totally determined by their kernels and ranges and the ranges are the same, the kernels have to converge. The rigorous proof takes a little bit of doing (cf. [S3], [S4]) ■

## 6.2 Irreducibility

A scheme (affine variety)  $\mathcal{Z}$  called irreducible if  $\mathcal{Z}$  is not a union of two proper subvarieties of  $\mathcal{Z}$ . So what does the irreducibility of  $\mathcal{B}_g$  has to do with Hermite projectors?

Before answering this question, let us point out a few facts. In what follows we will freely identify the projectors in  $\mathfrak{P}_G$  with points in  $\mathcal{B}_g$ .

**Proposition 6.4** *The set  $\mathcal{L}_g := \{P \in \mathcal{B}_g : P \text{ is Lagrange}\}$  is a Zariski open subset of  $\mathcal{B}_g$ , i.e., the complement to  $\mathcal{L}_g$  is a subvariety of  $\mathcal{B}_g$ . In particular, Zariski and Euclidean closure of  $\mathcal{L}_g$  in  $\mathcal{B}_g$  coincides and the dimension of this closure is  $dN$ .*

**Proof.** To prove that the complement of  $\mathcal{L}_g^c$  of  $\mathcal{L}_g$  is Zariski closed, we need to show that the points  $P \in \mathcal{L}_g^c$  are characterized by being solutions of polynomial equations. This can be easily done (cf. [S5]). The coefficients defining Lagrange projectors  $P \in \mathcal{L}_g$  are solutions to interpolation problems, hence, by Cramer’s rule can be expressed as ratios of determinants, that are polynomials in their entries (cf. [S5]). The entries depend on  $N$  interpolation points in  $\mathbb{C}^d$  (dimension  $Nd$ ). The upshot is that  $\mathcal{L}_g$  has a rational parametrization, hence the Zariski closure of  $\mathcal{L}_g$  is irreducible. Moreover, since  $\mathcal{L}_g$  is Zariski open in an irreducible variety  $\tilde{\mathcal{L}}_g$ , its Zariski closure coincides with its Euclidean closure (cf. [Mu]) which, by theorem 6.3, is the subvariety of  $\mathcal{B}_g$  that consists of Hermite projectors. ■



Let  $\mathcal{H}_{\mathfrak{g}} \subset \mathcal{B}_{\mathfrak{g}}$  be the irreducible subvariety of  $\mathcal{B}_{\mathfrak{g}}$  consisting of Hermite projectors. We can now answer the question posed at the beginning of this subsection:

**Theorem 6.5** ([S5]) *Let  $\mathfrak{g}$  be any basis for  $G$ . Then  $\mathcal{H}_{\mathfrak{g}}$  is irreducible subvariety of dimension  $\dim \mathcal{H}_{\mathfrak{g}} = dN$  and  $\mathcal{B}_{\mathfrak{g}}$  is irreducible if and only if every  $P \in \mathfrak{P}_G$  is Hermite. In this case  $\dim \mathcal{B}_{\mathfrak{g}} = dN$ .*

**Proof.** If  $\mathcal{H}_{\mathfrak{g}} \neq \mathcal{B}_{\mathfrak{g}}$  then, since the complement  $\mathcal{L}_{\mathfrak{g}}^c$  of  $\mathcal{L}_{\mathfrak{g}}$  is a subvariety of  $\mathcal{B}_{\mathfrak{g}}$ , it follows that  $\mathcal{B}_{\mathfrak{g}} = \mathcal{H}_{\mathfrak{g}} \cup \mathcal{L}_{\mathfrak{g}}^c$  and  $\mathcal{B}_{\mathfrak{g}}$  is reducible. ■

### 6.3 Hermite and non-Hermite projectors

Here is an immediate corollary of Theorem 6.3:

**Proposition 6.6** *Let  $P \in \mathfrak{P}_G$ . The following are equivalent:*

- (AT)  *$P$  is Hermite, i.e., can be approximated by Lagrange projectors*
- (PDE)  *$(\ker P)^\perp$  is the limit of subspaces consisting of pure exponentials*
- (LA)  *$\mathbf{M}_P$  can be approximated by a commuting sequence of diagonalizable matrices*

**Corollary 6.7** *If  $G \subset \mathbb{C}[x, y]$  then every  $P \in \mathfrak{P}_G$  is Hermite and the scheme  $\mathcal{B}_{\mathfrak{g}}$  is irreducible.*

**Proof.** By Proposition 6.6, we only need to prove that every pair of commuting matrices can be approximated by diagonalizable matrices, since diagonalizable matrices correspond to Lagrange projectors. But this is a known fact from [MT]. For a nice proof of it see [BS] and [Gu]. ■

**Theorem 6.8** *If  $d > 2$  then for some  $G \subset \mathbb{C}[x_1, \dots, x_d]$  the scheme  $\mathcal{B}_{\mathfrak{g}}$  is not irreducible, hence there are projectors  $P \in \mathfrak{P}_G$  that are not Hermite.*

**Proof.** (Essentially due to Iarrobino [Ia], cf. also [S5] and [S6]). Let  $U \subset \mathbb{C}[[\mathbf{x}]]$  be the span of ruffly half of monomials of degree  $n$  and let  $V$  be the other half. Now let  $E \subset \mathbb{C}[[\mathbf{x}]]$  be the space spanned by  $\mathbb{C}[\mathbf{x}]_{<n}$  and polynomials of the form

$$u - \sum_{v \in V} a_{v,u} v \quad (6.1)$$

for some coefficients  $a_{v,u}$ . Each such  $E$  is  $D$ -invariant, hence for each such  $E$  there exists  $P \in \mathfrak{P}_{\text{span}\{\mathbb{C}[\mathbf{x}]_{<n} \cup V\}}$  with  $(\ker P)^\perp = E$  we conclude that the dimension of  $\mathcal{B}_{\mathfrak{g}}$  is at least as large as the cardinality of the set  $\{(u, v) : u \in U, v \in V\}$ . Easy calculations show that, for  $d > 2$  and sufficiently large  $n$ , this cardinality is greater than  $d \times \dim(\text{span}\{\mathbb{C}[\mathbf{x}]_{<n} \cup V\})$  which, by Theorem 6.5 proves the corollary. ■

Compare this theorem to the next proposition to see how different  $\mathfrak{P}_G$  could be for different  $G$  of the same dimension:

**Proposition 6.9** *Every curvilinear projector is Hermite. Hence for every  $N$  and  $d$  there exists  $G \subset \mathbb{C}[\mathbf{x}]$  such that  $\mathcal{B}_{\mathfrak{g}}$  is irreducible.*

**Proof.** Let  $P$  be curvilinear and, without loss of generality the matrix  $M_1$  in  $\mathbf{M}_P$  be nonderogatory. Then there are polynomials  $p_i, i = 2 : d$  such that  $M_i = p_i(M_1)$ . Approximate  $M_1$  by diagonalizable matrices  $M_1^{(n)}$ . Then  $p_i(M_1^{(n)})$  are diagonalizable, approximate  $M_i$  and commute. Thus  $P$  is Hermite. Now for any  $N$  let  $G = \text{span}\{1, \dots, x_1^{N-1}\}$  considered as a subspace of  $\mathbb{C}[x_1, \dots, x_d]$ . Then, by definition, every  $P \in \mathfrak{P}_G$  is Hermite. ■

The last proposition appeared in [S3] with a much more difficult proof.

What about Gorenstein ideals?

**Theorem 6.10** (cf. [CN]) *In three variables the subvariety of Gorenstein ideals in  $\mathcal{B}_g$  is irreducible. In particular every projector  $P$  with Gorenstein kernel is Hermite.*

And this leads to a very sad personal story:

## 6.4 The tail of one projector

In [S6] the author gave many convincing reasons why the projector  $P_0$  from  $\mathbb{C}[x, y, z]$  onto its ten-dimensional subspace of quadratic polynomials  $G = \mathbb{C}[x, y, z]_{\leq 2}$ , given by

$$(\ker P_0)^\perp = \mathcal{D}(x^3 - yz, y^3 - xz, z^3 - xy) \quad (6.2)$$

is not Hermite. It would have been great if it wasn't. You see, from the argument in the Theorem 6.8 it follows that in three variable there exists a subspace  $G$  of dimension 102 such that the corresponding  $\mathcal{B}_g$  is not irreducible. It is now known (cf. [CEVV]) that every  $G \subset \mathbb{C}[x, y, z]$  with  $\dim G \leq 8$  induces an irreducible scheme  $\mathcal{B}_g$ . What happens between 102 and 8 is a wide open question. If my conjecture was true, it would have settle that question. Unfortunately it is not:

**Theorem 6.11** *The projector  $P_0$  is Hermite.*

**Proof.** Consider a projector  $P$  from  $\mathbb{C}[x, y, z]$  onto the span  $\{\mathbb{C}[x, y, z]_{\leq 2} \cup xyz\}$  given by

$$(\ker P)^\perp = \mathcal{D}\left(\frac{1}{4}(x^4 + y^4 + z^4) - xyz\right). \quad (6.3)$$

The ideal  $\ker P$  is Gorenstein and by the theorem 6.10,  $P$  is Hermite. By proposition 6.6 (PDE) this means that there exist coefficients  $(a_j^{(n)}, j = 1 : 11)$  and points  $\{(x_j^{(n)}, y_j^{(n)}, z_j^{(n)}) \rightarrow 0, j = 1 : 11\}$  such that

$$a_{11}^{(n)} e^{x_{11}^{(n)} x + y_{11}^{(n)} y + z_{11}^{(n)} z} + \sum_{j=1}^{10} a_j^{(n)} e^{x_j^{(n)} x + y_j^{(n)} y + z_j^{(n)} z} \rightarrow \frac{1}{4}(x^4 + y^4 + z^4) - xyz. \quad (6.4)$$

Multiplying both parts by  $e^{-(x_{11}^{(n)} x + y_{11}^{(n)} y + z_{11}^{(n)} z)}$  and letting

$$(\tilde{x}_j^{(n)}, \tilde{y}_j^{(n)}, \tilde{z}_j^{(n)}) := (x_j^{(n)}, y_j^{(n)}, z_j^{(n)}) - (x_{11}^{(n)}, y_{11}^{(n)}, z_{11}^{(n)}) \quad (6.5)$$

for  $j = 1, \dots, 10$  we have

$$a_{11}^{(n)} + \sum_{j=1}^{10} a_j^{(n)} e^{\tilde{x}_j^{(n)} x + \tilde{y}_j^{(n)} y + \tilde{z}_j^{(n)} z} \rightarrow \frac{1}{4}(x^4 + y^4 + z^4) - xyz. \quad (5.6)$$

Differentiating with respect to  $D_x$ ,  $D_y$  and  $D_z$  we conclude that, the Lagrange projectors  $P_n$  at ten points:

$$(\tilde{x}_j^{(n)}, \tilde{y}_j^{(n)}, \tilde{z}_j^{(n)}) \in \mathbb{C}^3 \quad (6.7)$$

have the property that for every  $F \in (\ker P_0)^\perp$  there exist  $F_n \in (\ker P_n)^\perp$  such that  $F_n \rightarrow F$ . Once again by proposition 6.6 (PDE), this proves the theorem.  $\blacksquare$

## 6.5 Smoothness

**Definition 6.12** *A scheme is smooth if at every point of the scheme the dimension of the tangent space coincides with the dimension of the scheme.*

To compute the dimension of the tangent space of  $\mathcal{B}_{\mathfrak{g}}$  one has to differentiate the equations in the definition of section ??, which is a rather tedious task. Fortunately there is a somewhat of a short cut.

**Theorem 6.13** (cf. [MS]) *For every ideal  $J \in \mathcal{B}_{\mathfrak{g}}$  the dimension of the tangent space to  $\mathcal{B}_{\mathfrak{g}}$  at  $J$  is equals to*

$$\dim \operatorname{Hom}_{\mathbb{C}[\mathbf{x}]}(J, \mathbb{C}[\mathbf{x}]/J). \quad (6.8)$$

Let me explain what does the term in (6.8) means.  $\operatorname{Hom}_{\mathbb{C}[\mathbf{x}]}(J, \mathbb{C}[\mathbf{x}]/J)$  is all mappings  $S$  from the ideal  $J$  into the quotient algebra  $\mathbb{C}[\mathbf{x}]/J$  such that

$$S(p_1 f_1 + p_2 f_2) = [p_1 S f_1] + [p_2 S f_2] \in \mathbb{C}[\mathbf{x}]/J \quad (6.9)$$

for all polynomials (!)  $p_1, p_2 \in \mathbb{C}[\mathbf{x}]$ .

For instance, let  $T$  be a Taylor projector from  $\mathbb{C}[x, y, z]$  onto span of  $\{1, x, y, z\}$ . Then  $J := \ker T$  is generated by monomials of degree 2 and

$$\mathbb{C}[\mathbf{x}]/J = \operatorname{span}\{[1], [x], [y], [z]\}.$$

Now imagine that  $S \in \operatorname{Hom}_{\mathbb{C}[\mathbf{x}]}(J, \mathbb{C}[\mathbf{x}]/J)$  is such that

$$\begin{aligned} S(x^2) &= a[1] + b[x] + c[y] + d[z] \\ S(xy) &= a_1[1] + b_1[x] + c_1[y] + d_1[z] \end{aligned}$$

Since  $x^2 y \in J$  and by (6.9) it follows that

$$\begin{aligned} S(y(x^2)) &= a[y] + b[xy] + c[y^2] + d[yz] = a[y] = \\ S(x(xy)) &= a_1[x] + b_1[x^2] + c_1[xy] + d_1[xz] = a_1[x] \end{aligned}$$

since all quadratic terms are in  $J$ . Thus  $a = a_1 = 0$  and the rest of the coefficients can be chosen arbitrary. In short, homomorphisms map every generator of  $J$  into an arbitrary linear combination of  $[x]$ ,  $[y]$  and  $[z]$ . Since there are 6 generators (quadratic monomials), the dimension of  $\text{Hom}_{\mathbb{C}[\mathbf{x}]}(J, \mathbb{C}[\mathbf{x}]/J) = 6 \times 3 = 18$ , but the dimension of  $\mathcal{B}_{\mathfrak{g}} = \dim G \times d = 4 \times 3 = 12$  and  $\mathcal{B}_{\mathfrak{g}}$  is not smooth.

Now let us look at another example:

The projector  $P$  is from  $\mathbb{K}[x, y, z, w]$  onto the 8-dimensional subspace  $G$  spanned by  $\mathfrak{g} := \{1, x, y, z, w, xz, yz, yw\}$  defined by  $Pxw = yz$  and  $Pu = 0$  for the rest of monomials not in  $G$ . Computing  $\text{Hom}_{\mathbb{C}[\mathbf{x}]}(J, \mathbb{C}[\mathbf{x}]/J)$  we have

$$\dim \text{Hom}_{\mathbb{C}[\mathbf{x}]}(J, \mathbb{C}[\mathbf{x}]/J) = 25 < \dim G \times 4 = 32$$

and this is interesting. The dimension of an irreducible algebraic variety cannot be less than the dimension of the tangent space at any point. Hence

**Theorem 6.14** ([EI]) *For  $\mathfrak{g} := \{1, x, y, z, w, xz, yz, yw\}$ , the border scheme is not irreducible and the projector  $P$  defined above is not Hermite.*

This is about the only explicitly known ideal projector that is not Hermite.

## 7 Deserts

Nothing tastes sweeter to a mathematician than an open problem. Here we will present a few of them.

### 7.1 Gorenstein Ideals

**Problem 7.1** *What is the relationship between the interpolating property of an ideal projector and the fact that  $\ker P$  is Gorenstein?*

The proof of Theorem 6.10 in [CN] is very involved.

**Problem 7.2** *Find a “linear algebra” proof of the following consequence of Theorem 6.10: Let  $(A, B, C)$  be a cyclic triple of commuting  $N \times N$  matrices such that  $(A^*, B^*, C^*)$  is also cyclic. Then there exist triples of commuting diagonalizable  $N \times N$  matrices  $(A_n, B_n, C_n)$  such that  $(A_n, B_n, C_n) \rightarrow (A, B, C)$ .*

### 7.2 Hermite projectors in three variables

**Problem 7.3** *What is the least  $N$  such that there exists an  $N$ -dimensional subspace  $G \subset \mathbb{C}[x, y, z]$  and a projector  $P \in \mathfrak{P}_G$  which is not Hermite?*

Such number must be between 9 and 102.

**Problem 7.4** *Find one non-Hermite projector on  $\mathbb{C}[x, y, z]$  explicitly.*

**Problem 7.5** (Guralnick [Gu]) *What is the least  $N$  such that there exists a triple  $(A, B, C)$  of commuting  $N \times N$  matrices that are not approximated by diagonalizable commuting matrices?*

That number must be between 9 and 30. Observe that the cyclicity assumption on  $(A, B, C)$  is omitted, which makes it a different problem from problem 7.3.

**Problem 7.6** *Find a triple  $(A, B, C)$  of commuting  $N \times N$  matrices explicitly that are not approximated by diagonalizable commuting matrices.*

The next problem is of philosophical nature:

**Problem 7.7** *What is so special about three(!) variables that make these questions so much more difficult than the similar questions in two or four variable? Is three variables somehow on the cusp?*

### 7.3 Connectedness

**Problem 7.8** *Is the border scheme  $\mathcal{B}_{\mathfrak{g}}$ , where  $\mathfrak{g}$  is the monomials basis for  $G$ , connected? In other words, given two ideal projector  $P_0, P_1 \in \mathfrak{P}_G$  does there exist a continuous family of ideal projectors  $P(t) \in \mathfrak{P}_G$  such that  $P(0) = P_0$  and  $P(1) = P_1$ ?*

For a monomial  $D$ -invariant space  $G \subset \mathbb{k}[\mathbf{x}]$  this problem was posed in [R2]:

Since every irreducible variety is connected, the answer is affirmative for  $G \subset C[x, y]$ . The answer is also affirmative for those  $D$ -invariant monomial spaces  $G$  where  $\deg f \geq \max\{\deg g : g \in G\}$  for every  $f \in \partial \mathfrak{g}$ . It is interesting to note that it follows from the general theory of Groebner basis (cf. [MS, Remark 18.3], [R2]), that “the Hilbert schemes” are connected. That is, for a given pair of ideal projectors  $P_0, P_1 \in \mathfrak{P}_G$  there exist a continuous family  $P(t)$  of ideal projectors, such that  $P(0) = P_0$  and  $P(1) = P_1$  and  $\dim \operatorname{ran} P = \dim G$  for all  $t$ . The rub is: for some  $t$  the ideal projector  $P(t)$  may project onto a subspace of dimension  $N$  that is different from  $G$ .

### 7.4 Analysis (finally)

Let  $P \in \mathfrak{P}_G$  be a Hermite projector. That means that there curves  $Z(t) = (\mathbf{z}_1(t), \dots, \mathbf{z}_N(t))$ ,  $\mathbf{z}_k(t) \in (\mathbb{C}^d)^N$  such that the Lagrange projector  $P(t) \in \mathfrak{P}_G$  interpolating at  $Z(t)$  converge to  $P$ . It is intuitively obvious that the trajectory of the curves  $\mathbf{z}_k(t)$  must “resemble” the surfaces  $b - Pb, b \in \partial \mathfrak{g}$ . In other words Given a Hermite projector  $P \in \mathfrak{H}_G$  and a family of Lagrange projectors  $P(t) \rightarrow P$ , what is the relationship between “trajectories” of  $\mathcal{Z}(t) = \mathcal{V}(\ker P(t))$  and  $P$ ?

**Problem 7.9** *Define “resemble”.*

**Illustration:** Recall the two primary ideal projectors  $T$  and  $P_*$  onto  $G = \mathbb{C}[x, y]_{\leq 1}$  given by (3.5). If all the points in  $\mathcal{Z}(t)$  tend to the origin along the curve  $(t, t^2) \subset \mathbb{C}^2$  then the corresponding Lagrange projectors  $P(t)$  converge to  $P_*$ . Computational evidence suggests that even a slight modification of this trajectory leads to a family of Lagrange projectors that converges to the Taylor projector  $T$ . In fact a “random” choice  $\mathcal{Z}(t) \rightarrow 0$  leads to a family of Lagrange projectors  $P(t)$  that converge to  $T$ . It seems that the Taylor projector is much more central in  $\mathfrak{P}_G$  than  $P_*$ . In other words, neighborhoods of  $T$  in  $\mathfrak{P}_G$  have larger density than neighborhoods of  $P_*$ . This suggests that, even in this simple case, the “geology” of  $\mathfrak{P}_G$  is not well-understood.

**Acknowledgement:** I would like to thank Carl de Boer for teaching me all about the subject and patiently coaching on its finer points. I also want to thank Martin Kreuzer and Lorenzo Robbiano for exposing me to the algebraic theory of the border schemes. I want to thank my student, Tom McKinley for pleasant discussions we have while smoking outside the department. My appreciations go to many anonymous referees from algebraic geometry whose friendly, and not so friendly, criticism taught me a lot. Finally I want to thank my friend George for his hospitality during this and many previous conferences.

## References

- [B] H. Bass, On the ubiquity of Gorenstein rings, *Math. Zeitschr.* 82, (1963), 8–28
- [Bi] Birkhoff, G., The algebra of multivariate interpolation, in *Constructive approaches to mathematical models*, C. V. Coffman and G. J. Fix (eds), Academic Press, New York, 1979, 345–363.
- [Bo1] Boor, C. de, Ideal interpolation, in *Approximation Theory XI, Gatlinburg 2004*, Chui, C. K., M. Neamtu and L. Schumaker (eds.), Nashboro Press (2005), 59–91.
- [Bo2] Boor, C. de, What are the limits of Lagrange projectors?, in *Constructive Theory of Functions, Varna 2005*, B. Bojanov, ed., Marin Drinov Academic Publishing House, Sofia, (2006), 51–63.
- [Bo3] Boor, C. de, Interpolation from spaces spanned by monomials, *Adv. Comput. Math.* 26 (1–3) (2007) 63–70.
- [BR] Boor, C. de, and A. Ron, On polynomial ideals of finite codimension with applications to box spline theory, *J. Math. Anal. Appl.* 158 (1991), 168–193.
- [BS] Boor, C. de, and B. Shekhtman, On the pointwise limits of bivariate Lagrange projectors, *LAA*, 429 (2008), 311–325.
- [CEVV] Cartwright, Dustin A., Daniel Erman, Mauricio Velasco, Bianca Viray, Hilbert schemes of 8 points in  $\mathbb{A}^d$ , arXiv:0803.0341.

- [CN] Casnati, G. and R., Notari, On the Gorenstein locus of some Punctual Hilbert Schemes, arXiv:0803.1135.
- [Co] Cox, D., Solving equations via algebras, in Solving Polynomial Equations, Foundations, Algorithms, and Applications, (A., Dickenstein and I., Z., Emiris eds.) Algorithms and Computation in Mathematics, 14, Springer 2005, 63–123.
- [CLO1] Cox, D., J. Little and D. O’Shea, Ideals, Varieties, and Algorithms, (second edition), Springer-Verlag, New-York-Berlin-Heidelberg, 1997.
- [CLO2] Cox, D., J. Little and D. O’Shea, Using Algebraic Geometry, Graduate Texts in Mathematics, Springer-Verlag, New-York-Berlin-Heidelberg, 1997.
- [EI] Iarrobino, A., and Emsalem, J., Some zero-dimensional generic singularities; Finite Algebras Having Small Tangent Space, *Compositio Mathematica*, 36(2), 1978, 145–188
- [Fo] Fogarty, J., (1968): Algebraic families on an algebraic surface. *Amer. J. Math.*, 90:511–521.
- [Ge] M. Gerstenhaber, On dominance and varieties of commuting matrices, *Annals of Mathematics* 73 (1961), 324–348.
- [Gu] Guralnick, R., A note on commuting pairs of matrices, *Linear and Multilinear Algebra* 31 (1992) 71–75.
- [GS] Guralnick, R., and B. Sethurman, Commuting pairs and triplets of matrices and related varieties, *Linear Algebra Appl.* 310 (2000) 139–148.
- [Ia] Iarrobino, A., Reducibility of the families of 0-dimensional schemes on a variety, *Inventiones Math.* 15, (1972), 72–77.
- [KR] M. Kreuzer and L. Robbiano, *Computational Commutative Algebra 2*, Springer, Heidelberg (2005).
- [Le] Lefranc, M., Analyse Spectrale sur  $\mathbb{Z}_n$ , *C. R. Acad. Sci. Paris* 246, (1958), 1951–1953
- [Ma] Macaulay, F. S., *The algebraic theory of modular systems*, Cambridge University Press, 1916 (reprinted 1994).
- [MSh] McKinley, T., and B. Shekhtman, On simultaneous block-diagonalization of cyclic sequences of commuting matrices, *Linear and Multilinear Algebra*, DOI: 10.1080/03081080802443125
- [MS] Miller E., and Sturmfels B., *Combinatorial Commutative Algebra*, Graduate Texts in Mathematics, 227, Springer 2000.

- [M] Möller, H. M., Hermite interpolation in several variables using ideal-theoretic methods, In *Constructive Theory of Functions of Several Variables*, W. Schempp and K. Zeller (eds.), Lecture Notes in Mathematics, Springer, (1977), 155–163.
- [Mo] B. Mourrain, A new criterion for normal form algorithms, in: Applied Algebra, in *Algebraic Algorithms and Error-Correcting Codes* (13th Intern. Symp., AAEEC-13, Honolulu, Hawaii USA, Nov.'99, Proc.), Mark Fossorier, Hideki Imai, Shu Lin, Alan Pol(eds.), Springer Lecture Notes in Computer Science, 1719, Springer-Verlag, Heidelberg, 1999, 430–443.
- [MT] Motzkin T., S., and, O. Taussky, Pairs of matrices with property L. II, Trans. Amer. Math. Soc. 80 (2) (1955) 387–401.
- [Mu] Mumford, D., The Red Book of Varieties and Schemes, Lecture Notes in Mathematics 1358, Springer-Verlag 1988.
- [N] H. Nakajima, Lectures on Hilbert schemes of points on surfaces, Amer. Math. Soc. University Lecture Series, vol.18, Providence RI, 1999.
- [R1] L. Robbiano, Zero-dimensional ideals or the inestimable value of estimable terms, in *Constructive Algebra and Systems Theory*, B. Hanzon, M. Hazewinkel (Eds.), Royal Netherlands Academy of Arts and Sciences, Netherlands, 2006, pp. 95–114.
- [R2] L. Robbiano, On border basis and Groebner basis schemes, arXiv:mathn0802.2793v2. To appear in *Collectanea Mathematica*, Vol 60(1).
- [S1] Shekhtman, B., On one Question of Ed Saff, Elec. Trans. Numer. Anal., Vol 25, (2006), 439—445.
- [S2] Shekhtman, B., On a conjecture of Carl de Boor regarding the limits of Lagrange Interpolants. Constr. Approx. 24(3), (2006), 365–370.
- [S3] Shekhtman, B., Bivariate ideal projectors and their perturbations, Adv. Comput. Math., Volume 29, Number 3, (2008), 207–228.
- [S4] Shekhtman, B., On perturbations of ideal complements, in: B. Randrianantonina, N. Randrianantonina (Eds.), *Banach Spaces and their Applications in Analysis*, De Gruyter, Berlin-New York, 2007, pp. 413–422.
- [S5] Shekhtman, B., On the limits of Lagrange projectors, *Constructive Approximation*, doi:10.1007/s00365-008-9016-0
- [S6] Shekhtman, B., Ideal Interpolation, Translations to and from Algebraic Geometry, Approximate Commutative Algebra, to appear in "Texts and Monographs in Symbolic Computation".



- [St] Stetter, H.J., Numerical Polynomial Algebra, SIAM, Philadelphia 2004.
- [Stu] Sturmfels, B., "Four counterexamples in combinatorial algebraic geometry", Journal of Algebra 230 (2000), p282–294.

## SEMIPARAMETRIC ESTIMATORS FOR HEAVY TAILED DISTRIBUTIONS

Valeria Caviezel  
Department MSIA  
University of Bergamo,  
Via dei Caniana, 2 - 24127-Bergamo- Italy  
E-mail: valeria.caviezel@unibg.it

Sergio Ortobelli  
Department MSIA  
University of Bergamo,  
Via dei Caniana, 2 - 24127-Bergamo- Italy  
E-mail: sergio.ortobelli@unibg.it

Prof. Svetlozar (Zari) T.Rachev  
Chair of Econometrics, Statistics  
and Mathematical Finance  
School of Economics and Business Engineering  
University of Karlsruhe and KIT,  
Kollegium am Schloss, Bau II, 20.12, R210  
Postfach 6980, D-76128, Karlsruhe, Germany and  
Department of Statistics and Applied Probability  
University of California, Santa Barbara  
CA 93106-3110, USA and  
Chief Scientist of FinAnalytica INC.  
E-mail: rachev@statistik.uni-karlsruhe.de

**Abstract:** In this paper we describe and apply estimating function theory to evaluate the parameters of parametric distributions uniquely defined by their characteristic functions. We first implement an estimating function model based on the first four moments of a parametric function of the underlying random variables. For instance we propose two parametric functions of the underlying random variables so as to obtain its first moments more easily by the simple knowledge of the characteristic function. Thus we consider the estimates that present the minimal asymptotic variance with respect to the parameter of the function. Then we propose an empirical analysis based on simulated stable Paretian distributions. Using simulated data of stable distributions we evaluate the forecasting power of the proposed methodology comparing it with the analogous maximum likelihood estimates. Moreover we show how to apply the same methodology to some well-known infinitely divisible distributions.

**Key words:** Estimating function, Stable distributions, infinitely divisible distributions, asymptotic variance.

## 1 Introduction

Several empirical investigations have shown that empirical distributions of many observed economic and financial data deviate from the ideal Gaussian law, since they often exhibit skewness and fat tails. Thus many alternative distributions have been proposed to approximate the underlying data. The classical example is the case of infinitely divisible distributions uniquely defined by their characteristic functions. Infinitely divisible distributions take into account the skewness and kurtosis of return series. In addition their associated Lévy processes can be used for valuing intertemporal financial products such as the derivatives. Thus several Lévy processes have been widely used in recent financial literature (see, among others, Rachev and Mitnik (2000) and the references therein). In particular, the stable Paretian distributions probably present the most attractive modeling properties, since they not only provide a better empirical fit, but they also possess heavy tails and result as limit distribution of sums of i.i.d. random variables.

In this paper, we deal with the problem of valuing the parameters of distributions uniquely defined by their characteristic function. Typically, if we know the characteristic function, we can determine the approximated distribution by inverting the characteristic function with the Fast Fourier Transform (FFT) and then we can value the maximum likelihood estimates (MLE) of the parameters. However, even if this methodology is widely used it does not permit an *a priori* valuation of the efficiency of the estimates. In this paper we propose a semiparametric methodology that gives optimal estimates of distribution parameters based on a valuation *a priori* of the asymptotic variance of the estimates. The semiparametric valuation is based on estimating function (EF) theory (see Godambe (1991) and the references therein). We suggest using this estimation either for the moments curve or for a bounded parametric function of the underlying distribution that admits finite the first four moments. Thus we propose to minimize the asymptotic variance of the estimates subject to some estimating equations. Finally we value the forecasting power of some of these semiparametric estimators applied to stable paretian distributions. In particular, we compare the absolute difference between simulated data, and either EF estimates or MLE ones obtained for stable paretian parameters. We also show how to use the same methodology for some particular infinitely divisible distributions.

The paper is organized as follows: Section 2 introduces semiparametric estimators based on estimating function theory, and in Section 3 we compare semiparametric and maximum likelihood methods. Finally, we briefly summarize the results.

## 2 Estimating Function parameter estimation

Suppose we have a sample  $X = (X_1, X_2, \dots, X_T)$  of i.i.d. observations whose distribution family  $\mathfrak{F}(\theta)$  is parametrized by  $\theta = (\theta_1, \theta_2, \dots, \theta_p) \in B \subseteq \mathbb{R}^p$ . In

the theory of estimating functions the optimum has two components: "unbiasedness" of EF and "smallness of the variance" of the standardized EFs. An estimating function  $h_i(X, \theta)$  is called *unbiased* if  $E(h_i(X, \theta)) = 0$  for all admissible  $\theta$ . In particular, when we consider a single parameter  $\theta$  the score function  $\frac{\partial \log f(X, \theta)}{\partial \theta}$  represents a the typical example of unbiased estimating function. We say that the unbiased estimating functions  $h_i(X_s, \theta)$  are *mutually orthogonal*, when  $E(h_i(X_s, \theta)h_j(X_s, \theta)) = 0$  for every  $i \neq j; i, j = 1, \dots, n$ . Among all the linear combinations  $l_{\theta, k} = \sum_{s=1}^T \sum_{i=1}^n a_{k,i}(\theta)h_i(X_s, \theta)$ , ( $k = 1, \dots, p$ ) of unbiased, mutually orthogonal estimating functions  $h_i(X_s, \theta)$ , the estimating functions  $l_{\theta, k}^* = \sum_{s=1}^T \sum_{i=1}^n a_{k,i}^*(\theta)h_i(X_s, \theta)$  with coefficients  $a_{k,i}^*(\theta) = E\left(\frac{\partial h_i}{\partial \theta_k}\right) / E(h_i^2)$ ,  $\forall i = 1, \dots, n; \forall k = 1, \dots, p$ , are optimal (with minimum variance). According to estimating function theory the optimal EFs  $\hat{\theta}_{(T)} = [\hat{\theta}_{(T),1}, \dots, \hat{\theta}_{(T),p}]$  obtained as consistent solution of the system:  $l_{\theta, k}^* = 0$ ,  $k = 1, \dots, p$ ; after orthogonalization, standardization and optimal combination is asymptotically Gaussian, i.e.,

$$\sqrt{T} \left( \hat{\theta}_{(T)} - \theta \right) \rightarrow MVN(0, V_{EF}^{-1}(\theta)) \text{ for } T \rightarrow \infty, \quad (1)$$

where  $V_{EF}(\theta) = [v_{i,j}]_{i,j=1,\dots,p}$  and  $v_{i,j} = E\left(\frac{\partial l_{\theta,i}^*}{\partial \theta_j}\right)$ ;  $i, j = 1, \dots, p$ . These results can be extended to stationary series that are not necessarily independent (see Li and Turtle (2000)). Moreover, we get similar results using GMM (General Method of Moments), but in this case we need a recursive methodology to determine the parameters in the linear combination  $l_{\theta,j}^*$ . Thus any algorithm based on GMM needs additional computational time.

Typical examples of optimal estimating functions are those proposed by Godambe and Thompson (1989) where they use two unbiased and mutually orthogonal estimating functions:  $h_1(X_s, \theta) = f(X_s) - m_f(\theta)$  and  $h_2(X_s, \theta) = (f(X_s) - m_f(\theta))^2 - \sigma_f^2(\theta) - s_f(\theta)\sigma_f(\theta)(f(X_s) - m_f(\theta))$ , where  $f$  is a measurable real function,  $m_f(\theta) = E(f(X_s))$ ,  $\sigma_f^2(\theta) = E((f(X_s))^2) - m_f^2(\theta)$ , and

$$s_f(\theta) = \frac{E((f(X_s) - m_f(\theta))^3)}{\sigma_f^3(\theta)}.$$

Therefore, we get the EF optimal estimator  $\hat{\theta}_{(T)}$  of the vector of parameters  $\theta$  solving the equations (for  $k = 1, \dots, p$ )

$$l_{\theta, k}^* = \sum_{s=1}^T (a_{k,1}h_1(X_s, \theta) + a_{k,2}h_2(X_s, \theta)) = 0, \text{ where}$$

$$a_{k,1} = \frac{E\left(\frac{\partial h_1(X_s, \theta)}{\partial \theta_k}\right)}{E(h_1^2(X_s, \theta))} = \frac{-\frac{\partial m_f(\theta)}{\partial \theta_k}}{\sigma_f^2(\theta)}, \quad (2)$$

$$a_{k,2} = \frac{E\left(\frac{\partial h_2(X_s, \theta)}{\partial \theta_k}\right)}{E(h_2^2(X_s, \theta))} = \frac{-\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k} + \frac{\partial m_f(\theta)}{\partial \theta_k} s_f(\theta) \sigma_f(\theta)}{\sigma_f^4(\theta) (k_f(\theta) - 1 - s_f^2(\theta))}, \quad (3)$$

and  $k_f(\theta) = \frac{E\left((f(X_s) - m_f(\theta))^4\right)}{\sigma_f^4(\theta)}$ . We call this class of estimating functions *GT (Godambe and Thompson) estimating functions*. Note that we can also easily obtain an analytical formulation of the asymptotic variance of GT estimating functions when we know  $\frac{\partial m_f(\theta)}{\partial \theta_k}$  and  $\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k}$ . As a matter of fact, from (1) we deduce that the asymptotic variance  $V_{EF}^{-1}(\theta)$  is the inverse of  $V_{EF}(\theta) = [v_{i,j}(\theta)]_{i,j=1,\dots,p}$ , where  $v_{i,j}(\theta) = E\left(\frac{\partial l_{\theta,i}^*}{\partial \theta_j}\right)$ . Thus, we get

$$v_{i,j}(\theta) = T \frac{\frac{\partial m_f(\theta)}{\partial \theta_i} \frac{\partial m_f(\theta)}{\partial \theta_j}}{\sigma_f^2(\theta)} + T \frac{\left(-\frac{\partial \sigma_f^2(\theta)}{\partial \theta_j} + \frac{\partial m_f(\theta)}{\partial \theta_j} s_f(\theta) \sigma_f(\theta)\right) \left(-\frac{\partial \sigma_f^2(\theta)}{\partial \theta_i} + \frac{\partial m_f(\theta)}{\partial \theta_i} s_f(\theta) \sigma_f(\theta)\right)}{\sigma_f^4(\theta) (k_f(\theta) - 1 - s_f^2(\theta))}. \quad (4)$$

It has been demonstrated (see Ortobelli and Topaloglou (2008)) that the solutions of GT estimating functions  $l_{\theta,k}^* = 0$  are given by the values  $\theta_k$ , solutions of the simplified equations for  $k = 1, \dots, p$ :

$$m_f(\theta) = \begin{cases} \frac{1}{T} \sum_{s=1}^T f(X_s) + c_k - b_k & \text{if } c_k > 0 \\ \frac{1}{T} \sum_{s=1}^T f(X_s) + c_k + b_k & \text{if } c_k \leq 0 \end{cases} \quad (5)$$

where

$$b_k = \left( c_k^2 - \frac{1}{T} \sum_{s=1}^T \left( f(X_s) - \frac{1}{T} \sum_{s=1}^T f(X_s) \right)^2 + \sigma_f^2(\theta) \right)^{1/2}, \text{ and } c_k = \frac{a_{k,1} - a_{k,2} s_f(\theta) \sigma_f(\theta)}{2a_{k,2}}.$$

In particular, when  $m_f(\theta)$  is itself a possible parameter, then a consistent estimating function of  $m_f(\theta)$  is given by (5) where  $c_k = \frac{a_{m,1} - a_{m,2} s_f(\theta) \sigma_f(\theta)}{2a_{m,2}}$ ,

$$a_{m,1} = \frac{-1}{\sigma_f^2(\theta)}, \quad a_{m,2} = \frac{-\frac{\partial \sigma_f^2(\theta)}{\partial m} + s_f(\theta) \sigma_f(\theta)}{\sigma_f^4(\theta) (k_f(\theta) - 1 - s_f^2(\theta))} \text{ and } \frac{\partial \sigma_f^2(\theta)}{\partial m_f} = \frac{\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k}}{\frac{\partial m_f(\theta)}{\partial \theta_k}} \text{ for a}$$

given  $k$ . According to formulas (2) (3), the above estimating equations (5) are uniquely determined by  $m_f(\theta)$ ,  $\sigma_f^2(\theta)$ ,  $s_f(\theta)$ ,  $k_f(\theta)$ ,  $\frac{\partial m_f(\theta)}{\partial \theta_k}$  and  $\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k}$  that are respectively the mean, the variance, the skewness and the kurtosis of  $f(X)$ , and the partial derivatives of  $m_f(\theta)$ , and  $\sigma_f^2(\theta)$ . Thus, if  $X$  admits finite the first four moments, we can use the function  $f(X) = X$  to get some estimators of the distributional parameters. Otherwise, we can use a bounded function  $f$ .

Moreover, in several cases it could be useful to adopt a parametric differentiable function  $f : A \rightarrow \mathbb{R}$ , where  $A = \mathbb{R} \times [a, b] \subseteq \mathbb{R}^2$ . As a matter of fact, in this case the previous estimators and functionals (i.e.,  $m_f(\theta, q)$ ,  $\sigma_f^2(\theta, q)$ ,  $s_f(\theta, q)$  and  $k_f(\theta, q)$ ) also depend on a parameter  $q \in [a, b]$  of the function  $f(X, q)$ . This aspect can be used to get an optimal estimator with the minimum asymptotic variance as suggested in the following remark.

**Remark** Suppose we have a function  $f : A \rightarrow \mathbb{R}$  (with  $A \subseteq \mathbb{R}^2$ ) such that for any real random variable  $X \in \mathfrak{S}(\theta)$  (with  $\theta \in B \subseteq \mathbb{R}^p$ )  $f(X, q)$  admits finite the first four moments for any admissible  $q$ . Then we get an optimal GT estimator with minimum asymptotic variance solving the following optimization problem:

$$\max_{(\theta, q) \in A} |V_{EF}(\theta, q)| \text{ subject to} \quad (6)$$

$$(5), \quad k = 1, \dots, p, \quad \theta \in B;$$

where  $|V_{EF}(\theta, q)|$  is the determinant of the inverse asymptotic variance (1)  $V_{EF}(\theta, q) = [v_{i,j}(\theta, q)]_{i,j=1,\dots,p}$  and  $v_{i,j}(\theta, q) = E\left(\frac{\partial l_{\theta,i}^*}{\partial \theta_j}\right)$ . In order to reduce the computational complexity of (6) we can use  $\sum_{i=1}^p v_{i,i}(\theta, q)$  as objective function instead of  $|V_{EF}(\theta, q)|$ .

Clearly the idea of minimizing the asymptotic variance subject to the constraints of some equations can be applied to many other estimators (for example the maximum likelihood estimator). Using a GT estimator we do not necessarily need to know a closed form of the density (or the cumulative) distribution of the underlying random variables. As a matter of fact, GT estimators can be used even for those distributions uniquely defined by their characteristic function  $\phi_X(t) = E(\exp(itX))$ . Let us consider two possible examples.

**Moment estimator:** Suppose all parametric random variables belong to the space  $L^r = \{X/E(|X|^r) < \infty\}$ . By using the derivatives of the characteristic function  $\phi_X^{(k)}(0)$  we can determine all the existing integer moments of  $X$ , since  $\phi_X^{(k)}(0) = i^k E(X^k)$ . Then, as parametric functions we can use:

- $f(X, q) = |X|^q$  for any  $q \in [0, r/4]$  (in this case  $m_f(\theta, q) = E(|X|^q)$  represents the moments curve that characterizes the distribution of  $|X|$ );
- $f(X, q) = X^q$  for any integer  $q \in [0, r/4]$ .

**Trigonometric estimator:** Suppose  $f(X, t) = \sin(tX)$  for some given  $t \neq 0$ . Then if we know the characteristic function  $\phi_X(t) = E(\exp(itX))$  we can easily determine the first four moments of  $f(X, t)$ , since  $E(\cos(tX)) = \text{Re}(\phi_X(t))$ ;  $E(\sin(tX)) = \text{Im}(\phi_X(t))$ . So, if  $f(X, t) = \sin(tX)$  for a  $t \neq 0$  the first four moments of  $\sin(tX)$  are given by:

$$\begin{aligned} E(\sin(tX)) &= \text{Im}(\phi_X(t)); \quad E(\sin^2(tX)) = 0.5(1 - \text{Re}(\phi_X(2t))) \\ E(\sin^3(tX)) &= \frac{1}{4}(3 \text{Im}(\phi_X(t)) - \text{Im}(\phi_X(3t))); \\ E(\sin^4(tX)) &= \frac{1}{8}(3 - 4 \text{Re}(\phi_X(2t)) + \text{Re}(\phi_X(4t))); \end{aligned} \quad (7)$$

and together with these we have to know  $\frac{\partial m_f(\theta, t)}{\partial \theta_k} = \frac{\partial \text{Im}(\phi_X(t))}{\partial \theta_k}$  and  $\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k} = \frac{-0.5 \partial \text{Re}(\phi_X(2t))}{\partial \theta_k} - 2 \text{Im}(\phi_X(t)) \frac{\partial \text{Im}(\phi_X(t))}{\partial \theta_k}$ . This method can be easily applied to estimate the parameters of all distributions defined by their characteristic function, when from the characteristic function we can easily distinguish the imaginary part from the real one. For example, all the infinitely divisible random variables  $X$  have characteristic function  $\phi_X(u)$  uniquely determined by the triplet  $[\gamma, \sigma^2, \nu]$  that identifies the so called Lévy-Khintchine characteristic exponent  $\psi_X(u) = \log \phi_X(u)$  given by:

$$\begin{aligned} \psi_X(u) &= i\gamma u - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{+\infty} (\exp(iux) - 1 - iux1_{\{|x|<1\}})\nu(dx) = \\ &= i\left(\gamma u + \int_{-\infty}^{+\infty} (\sin(ux) - ux1_{\{|x|<1\}})\nu(dx)\right) - \\ &\quad - \frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{+\infty} (\cos(ux) - 1)\nu(dx) \end{aligned}$$

where  $\gamma \in \mathbb{R}$ ,  $\sigma^2 > 0$  and  $\nu$  is a measure on  $\mathbb{R} \setminus \{0\}$  with  $\int_{-\infty}^{+\infty} (1 \wedge x^2)\nu(dx) < \infty$ . Therefore for any infinitely divisible random variables  $X$  we can easily identify the real and the imaginary one of the characteristic function given by:

$$\begin{aligned} \text{Re}(\phi_X(u)) &= \exp\left(-\frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{+\infty} (\cos(ux) - 1)\nu(dx)\right) \times \\ &\quad \times \cos\left(\gamma u + \int_{-\infty}^{+\infty} (\sin(ux) - ux1_{\{|x|<1\}})\nu(dx)\right), \\ \text{Im}(\phi_X(u)) &= \exp\left(-\frac{1}{2}\sigma^2 u^2 + \int_{-\infty}^{+\infty} (\cos(ux) - 1)\nu(dx)\right) \times \\ &\quad \times \sin\left(\gamma u + \int_{-\infty}^{+\infty} (\sin(ux) - ux1_{\{|x|<1\}})\nu(dx)\right). \end{aligned}$$

In particular, the Lévy triplet  $[\gamma, \sigma^2, \nu]$  identifies the three main components of any Lévy process: the deterministic component ( $\gamma$ ), the Brownian component ( $\sigma^2$ ) and the pure jump component ( $\nu$ ). For further details on theoretical aspects we refer to Sato (1999).

## 2.1 Examples of infinitely divisible distributions

In this subsection we consider some particular infinitely divisible distributions: stable Paretian distributions, tempered stable (TS) distributions, normal inverse Gaussian (NIG) distributions, and Carr, Geman, Madan, Yor (CGMY) distributions. For each of these distributions we describe:

- the characteristic function distinguishing the real and imaginary parts;
- the moments curve in the stable Paretian case and the mean, the variance, the skewness and the kurtosis for the other infinitely divisible distributions;

- in Table I we give the derivatives  $\frac{\partial m_f(\theta)}{\partial \theta_k}$  and  $\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k}$  for the GT (EF) moment estimator;
- in Tables II and III we give the derivatives  $\frac{\partial \text{Im}(\phi_X(t))}{\partial \theta_k}$  and  $\frac{\partial \text{Re}(\phi_X(t))}{\partial \theta_k}$  for the GT (EF) trigonometric estimator.

Doing so, we are able to derive:

- a GT trigonometric estimator when  $f(X, t) = \sin(tX)$  for all these distributions;
- a GT moment estimator when  $f(X) = X$  for TS, NIG and CGMY distributions;
- a GT parametric moment estimator when  $f(X, p) = |X - \mu|^p$  for stable Paretian distributions.

**Stable distributions** (see, among others, Rachev and Mittnik (2000) and the references therein) A univariate stable distribution  $X \stackrel{d}{=} S_\alpha(\sigma, \beta, \mu)$  is characterized by four parameters. These are: the index of stability  $\alpha \in (0, 2]$ , the scale parameter  $\sigma \in \mathbb{R}^+$ , the skewness parameter  $\beta \in [-1, +1]$  and the shift parameter  $\mu \in \mathbb{R}$ . The stable distribution is Gaussian when  $\alpha = 2$ , and in this case,  $\sigma$  is proportional to the standard deviation,  $\beta$  can be taken to be zero and  $\mu$  is the mean. An  $\alpha$  stable non Gaussian distribution admits finite  $p$ -th fractional moment  $E(|X|^p) < \infty$  for any  $p \in (-1, \alpha)$ . A stable distribution can be defined in different equivalent ways. We use the characteristic function extensively because few stable density functions are known in closed form. The characteristic function of  $X \stackrel{d}{=} S_\alpha(\sigma, \beta, \mu)$  is given by:

$$\phi_X(t) = \begin{cases} \exp\left(-|t\sigma|^\alpha \left(1 - i \operatorname{sgn}(t)\beta \tan \frac{\pi\alpha}{2}\right) + it\mu\right) & \text{if } \alpha \in (0, 2]; \alpha \neq 1 \\ \exp\left(-|t\sigma| \left(1 + i\beta \frac{2}{\pi} \operatorname{sgn}(t) \log |t|\right) + it\mu\right) & \text{if } \alpha = 1. \end{cases}$$

Thus,  $\operatorname{Re}(\phi_X(t)) = \exp(-|t\sigma|^\alpha) \cos(b)$ ;  $\operatorname{Im}(\phi_X(t)) = \exp(-|t\sigma|^\alpha) \sin(b)$ , where

$$b = \begin{cases} |t\sigma|^\alpha \operatorname{sgn}(t)\beta \tan \frac{\pi\alpha}{2} + t\mu & \text{if } \alpha \in (0, 2]; \alpha \neq 1 \\ t\mu - |t\sigma| \beta \frac{2}{\pi} \operatorname{sgn}(t) \log |t| & \text{if } \alpha = 1, \end{cases}$$

and we can apply formula (5) to approximate  $\alpha, \sigma, \beta, \mu$  with the GT trigonometric estimator when we assume  $f(X, t) = \sin(tX)$ . Thus using formulas (2) (3) (7) and (4) we can solve the problem (6):

$$\begin{aligned} & \max_{\alpha, \sigma, \beta, \mu, t} |V_{EF}(\alpha, \sigma, \beta, \mu, t)| \\ & \text{subject to (5); } t \neq 0, \end{aligned}$$



**Table 1**

This table summarizes the derivatives used in the moment estimators for TS, NIG, CGMY and Stable distributions.

$X = TS(k, a, b)$	
$\frac{\partial E(X)}{\partial k} = 2ab^{(k-1)/k} \left( 1 + \frac{1}{k} \ln b \right)$	$\frac{\partial Var(X)}{\partial k} = 4ab^{(k-2)/k} \frac{k + 2(\ln b - k^2 - k \ln b)}{k}$
$\frac{\partial E(X)}{\partial a} = 2kb^{(k-1)/k}$	$\frac{\partial Var(X)}{\partial a} = 4k(1-k)b^{(k-2)/k}$
$\frac{\partial E(X)}{\partial b} = 2ab^{-1/k}(k-1)$	$\frac{\partial Var(X)}{\partial b} = 4a(1-k)(k-2)b^{-2/k}$
$X = NIG(\alpha, \beta, \delta)$	
$\frac{\partial E(X)}{\partial \delta} = \beta(\alpha^2 - \beta^2)^{-1/2}$	$\frac{\partial Var(X)}{\partial \delta} = \alpha^2(\alpha^2 - \beta^2)^{-3/2}$
$\frac{\partial E(X)}{\partial \beta} = \alpha^2\delta(\alpha^2 - \beta^2)^{-3/2}$	$\frac{\partial Var(X)}{\partial \beta} = 3\alpha^2\beta\delta(\alpha^2 - \beta^2)^{-5/2}$
$\frac{\partial E(X)}{\partial \alpha} = -\alpha\beta\delta(\alpha^2 - \beta^2)^{-3/2}$	$\frac{\partial Var(X)}{\partial \alpha} = -\alpha\delta(\alpha^2 + 2\beta^2)(\alpha^2 - \beta^2)^{-5/2}$
$X = CGMY(C, G, M, Y)$	
$\frac{\partial E(X)}{\partial C} = (M^{Y-1} - G^{Y-1})\Gamma(1-Y)$	$\frac{\partial Var(X)}{\partial C} = (M^{Y-2} + G^{Y-2})\Gamma(2-Y)$
$\frac{\partial E(X)}{\partial G} = CG^{Y-2}\Gamma(2-Y)$	$\frac{\partial Var(X)}{\partial G} = -CG^{Y-3}\Gamma(3-Y)$
$\frac{\partial E(X)}{\partial M} = -CM^{Y-2}\Gamma(2-Y)$	$\frac{\partial Var(X)}{\partial M} = -CM^{Y-3}\Gamma(3-Y)$
$\frac{\partial E(X)}{\partial Y} = C\Gamma(1-Y)[(M^{Y-1}\ln M - G^{Y-1}\ln G) - \Psi(1-Y)(M^{Y-1} - G^{Y-1})]$	$\frac{\partial Var(X)}{\partial Y} = C\Gamma(2-Y)[(M^{Y-2}\ln M + G^{Y-2}\ln G) - \Psi(2-Y)(M^{Y-2} + G^{Y-2})]$
$X = S_\alpha(\beta, \sigma, \mu)$	
$\frac{\partial E( X ^p)}{\partial \sigma} = A(p) p \sigma^{p-1} (1 + \beta^2 r^2)^{p/(2\alpha)} \cos y$ $\frac{\partial E( X ^p)}{\partial \beta} = A(p) \sigma^p \frac{psr[0.5\beta r(1 + \beta^2 r^2)(2\cos y + (w + v)\cos \alpha^{-1}) - (1 + \beta^2)\sin y + 0.5(w - v)(1 - \beta^2)\sin \alpha^{-1}]}{\alpha(1 + \beta^2 r^2)(1 + q + \beta^2 - q\beta^2)}$ $\frac{\partial E( X ^p)}{\partial \alpha} = \frac{(2\sigma)^p}{\alpha\sqrt{\pi}} \Gamma\left(\frac{\alpha - p}{\alpha}\right) (1 + \beta^2 r^2)^{p/(2\alpha)} \left[ \left( y - \frac{\pi p \beta (1 + r^2)}{2(1 + \beta^2 r^2)} \right) \sin y + \frac{p}{\alpha} \Psi\left(\frac{\alpha - p}{\alpha}\right) \cos y + \frac{p}{2} \left( \frac{\pi p \beta^2 (1 + r^2)}{1 + \beta^2 r^2} - \frac{\ln(1 + \beta^2 r^2)}{\alpha} \right) \right]$	
<p>where: <math>q = \cos \pi \alpha</math>      <math>r = \tan \frac{\pi \alpha}{2}</math>      <math>A(p) = \frac{2^p \Gamma\left(\frac{\alpha - p}{\alpha}\right) \Gamma\left(\frac{p + 1}{2}\right)}{\sqrt{\pi} \Gamma\left(\frac{2 - p}{p}\right)}</math>      <math>s = \left( \frac{q + \beta^2 - q\beta^2 + 1}{2 \cos^2 \frac{\pi \alpha}{2}} \right)^{p/(2\alpha)}</math></p>	
$v = \pi \alpha^2 + p \arctan(\beta r)$ $w = \pi \alpha^2 - p \arctan(\beta r)$ $y = \frac{p}{\alpha} \arctan(\beta r)$	

**Table II**

This table summarizes the derivatives used in the trigonometric estimators for TS, and CGMY distributions.

$CGMY(C, G, M, Y)$	where
$\frac{\partial \text{Re}}{\partial C} = n\Gamma(-Y) \left[ r \cos(p+v-w) - \cos(v-w) (M^Y + G^Y) + s \cos(q-v+w) \right]$	$p = Y \arctg \frac{t}{G}$
$\frac{\partial \text{Re}}{\partial G} = \frac{nCY\Gamma(-Y)}{G^2 + t^2} \left[ tr \sin(p+v-w) - \cos(v-w) (t^2 G^{Y-1} + G^{Y+1}) + rG \cos(p+v-w) \right]$	$q = Y \arctg \frac{t}{M}$
$\frac{\partial \text{Re}}{\partial M} = \frac{nCY\Gamma(-Y)}{M^2 + t^2} \left[ ts \sin(q-v+w) - \cos(v-w) (t^2 M^{Y-1} + M^{Y+1}) + sM \cos(q-v+w) \right]$	$r = \sqrt{(G^2 + t^2)^Y}$
$\frac{\partial \text{Re}}{\partial Y} = nC\Gamma(-Y) \left\{ \cos(v-w) \left[ \Psi(-Y) (G^Y + M^Y) - M^Y \ln M - G^Y \ln G \right] - r \cos(p+v-w) \right.$ $\left. \left[ \Psi(-Y) - \frac{1}{Y} \ln r \right] - s \cos(q-v+w) \left[ \Psi(-Y) - \frac{1}{Y} \ln s \right] - \frac{tp}{Y} \cos(p+v-w) - \frac{sq}{Y} \cos(q-v+w) \right\}$	$s = \sqrt{(M^2 + t^2)^Y}$ $v = rC\Gamma(-Y) \sin p$
$\frac{\partial \text{Im}}{\partial C} = n\Gamma(-Y) \left[ r \sin(p+v-w) - \sin(v-w) (M^Y + G^Y) - s \sin(q-v+w) \right]$	$w = sC\Gamma(-Y) \sin q$
$\frac{\partial \text{Im}}{\partial G} = \frac{nCY\Gamma(-Y)}{G^2 + t^2} \left[ -tr \cos(p+v-w) - \sin(v-w) (t^2 G^{Y-1} + G^{Y+1}) + rG \sin(p+v-w) \right]$	$n = e^{C\Gamma(-Y) [r \cos p - M^t - G^t + s \sin q]}$
$\frac{\partial \text{Im}}{\partial M} = \frac{nCY\Gamma(-Y)}{M^2 + t^2} \left[ ts \cos(q-v+w) - \sin(v-w) (t^2 M^{Y-1} + M^{Y+1}) - sM \sin(q-v+w) \right]$	
$\frac{\partial \text{Im}}{\partial Y} = nC\Gamma(-Y) \left\{ \sin(v-w) \left[ \Psi(-Y) (G^Y + M^Y) - M^Y \ln M - G^Y \ln G \right] - r \sin(p+v-w) \right.$ $\left. \left[ \Psi(-Y) - \frac{1}{Y} \ln r \right] + s \sin(q-v+w) \left[ \Psi(-Y) - \frac{1}{Y} \ln s \right] + \frac{tp}{Y} \cos(p+v-w) - \frac{sq}{Y} \cos(q-v+w) \right\}$	
$TS(k, a, b)$	where
$\frac{\partial \text{Re}}{\partial a} = bw \cos(ay \sin z) - yw \cos p$	$z = k \arctan(2tb^{-1/k})$
$\frac{\partial \text{Re}}{\partial b} = \frac{aw}{b^{(k+3)/k} + 4t^2 b^{(k+1)/k}} \left[ \cos(ay \sin z) (b^{(k+3)/k} + 4t^2 b^{(k+1)/k}) - b^{3/k} y \cos p - 2tb^{2/k} y \sin p \right]$	$y = \sqrt{(b^{2/k} + 4t^2)^k}$
$\frac{\partial \text{Re}}{\partial k} = \frac{ayw}{kb^{2/k} + 4kt^2} \left[ \cos p (-b^{2/k} \ln y - 4t^2 \ln y + b^{2/k} \ln b) + \sin p (2tb^{1/k} \ln b + zb^{2/k} + 4zt^2) \right]$	$p = z - ay \sin z$
$\frac{\partial \text{Im}}{\partial a} = bw \sin(ay \sin z) + yw \sin p$	$w = \exp\{a(b - y \cos z)\}$
$\frac{\partial \text{Im}}{\partial b} = \frac{aw}{b^{(k+3)/k} + 4t^2 b^{(k+1)/k}} \left[ \sin(ay \sin z) (b^{(k+3)/k} + 4t^2 b^{(k+1)/k}) + b^{3/k} y \sin p - 2tb^{2/k} y \cos p \right]$	
$\frac{\partial \text{Im}}{\partial k} = \frac{ayw}{kb^{2/k} + 4kt^2} \left[ \sin p (-b^{2/k} \ln y + 4t^2 \ln y - b^{2/k} \ln b) + \cos p (2tb^{1/k} \ln b + zb^{2/k} + 4zt^2) \right]$	

where  $\frac{\partial \text{Im}(\phi_X(t))}{\partial \theta_k}$  and  $\frac{\partial \text{Re}(\phi_X(t))}{\partial \theta_k}$  are given in Table III. Moreover, we can propose a GT moment estimator based on the moments curve. As a matter of fact the absolute central moments for any  $p \in (-1, \alpha)$ ,  $\alpha \neq 1$  are given by:

$$E(|X - \mu|^p) = \sigma^p \left(1 + \beta^2 \tan^2 \frac{\alpha\pi}{2}\right)^{0.5p/\alpha} \cos\left(\frac{p}{\alpha} \arctan\left(\beta \tan \frac{\alpha\pi}{2}\right)\right) A(p),$$

$$\text{where } A(p) = \frac{2^p \Gamma\left(\frac{\alpha-p}{\alpha}\right) \Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{2-p}{2}\right) \sqrt{\pi}} \text{ and } \Gamma(c) = \int_0^{+\infty} z^{c-1} e^{-z} dz \text{ for } c > 0$$

while we refer to Abramowitz and Stegun's definition (see Abramowitz and Stegun 1970) of the Gamma function  $\Gamma(\cdot)$  of a negative real (non integer) number. Thus, if we use  $f(X) = |X - \mu|^p$  such that  $p$  is small enough (i.e.,  $p \in (-0.25, \alpha/4)$ ) we can easily get the first four moments  $E(f(X)^j) = E(|X - \mu|^{jp})$   $j = 1, \dots, 4$ , in order to obtain a GT moment estimator for the stable parameters. Thus we can apply formulas (2) (3) (using  $\frac{\partial E(|X-\mu|^p)}{\partial \alpha}$ ,  $\frac{\partial E(|X-\mu|^p)}{\partial \sigma}$  and  $\frac{\partial E(|X-\mu|^p)}{\partial \beta}$  given in Table I) to determine estimating functions (5) for  $\alpha, \sigma, \beta$ . Since we need four equations for four parameters and we do not know a closed form of  $E(|X - \mu|^p \text{sgn}(X - \mu))$ , we suggest using, as fourth estimating equation, the consistent estimating function of  $m_f(\theta)$  (5) assuming that  $m_f(\theta)$  is itself a possible parameter. Since  $\sigma_f^2(\alpha, \sigma, \beta) = E(|X - \mu|^{2p}) - m_f(\alpha, \sigma, \beta)^2$ ,

then  $\frac{\partial \sigma_f^2(\alpha, \sigma, \beta)}{\partial m_f} = \frac{\frac{\partial \sigma_f^2(\alpha, \sigma, \beta)}{\partial \sigma}}{\frac{\partial m_f(\alpha, \sigma, \beta)}{\partial \sigma}}$ ; and  $\frac{\partial m_f(\alpha, \sigma, \beta)}{\partial m_f} = 1$ , and the fourth estimating

equation is  $l_m^* = \sum_{s=1}^T (a_1 h_1(X_s, \theta) + a_2 h_2(X_s, \theta)) = 0$  where  $a_1 = \frac{-1}{\sigma_f^2(\alpha, \sigma, \beta)}$ ,

$$a_2 = \frac{-\frac{\partial \sigma_f^2(\theta)}{\partial m_f} + s_f(\theta) \sigma_f(\theta)}{\sigma_f^4(\theta) (k_f(\theta) - 1 - s_f^2(\theta))}.$$

Analogously we get the elements of the asymptotic variance (4). Doing so, we can maximize the inverse of the asymptotic variance determinant  $\max_{\alpha, \sigma, \beta, \mu, p} |V_{EF}(\alpha, \sigma, \beta, \mu, p)|$  associated with the GT moment estimator of stable paretian distributions.

**Tempered Stable (TS) distribution** (see Tweedie (1984) and, for more general definitions, see Kim *et al.* (2008)) Tempered stable distributions depend on three parameters  $a > 0$ ;  $b \geq 0$ ;  $0 < k < 1$  and their characteristic function is given by:

$$\phi_{TS}(u; k, a, b) = \exp \left( ab - a \left( b^{1/k} - 2iu \right)^k \right).$$

Thus,  $\text{Re}(\phi_X(t)) = c \cos(d)$ ;  $\text{Im}(\phi_X(t)) = c \sin(d)$ , where

$$c = \exp \left( ab - a \left( b^{2/k} + 4t^2 \right)^{k/2} \cos \left( k \arctan \frac{-2t}{b^{1/k}} \right) \right);$$

$$d = -a \left( b^{2/k} + 4t^2 \right)^{k/2} \sin \left( k \arctan \frac{-2t}{b^{1/k}} \right).$$

Moreover, if we assume  $f(X) = X$ , the mean, the variance, the skewness and the kurtosis of tempered stable distributions are given by:

$$\begin{aligned} m_f(a, b, k) &= 2akb^{(k-1)/k}; & \sigma_f^2(a, b, k) &= 4ak(1-k)b^{(k-2)/k}; \\ s_f(a, b, k) &= \frac{(k-2)}{[abk(1-k)]^{1/2}}; & k_f(a, b, k) &= 3 + \frac{4k-6-k(1-k)}{abk(1-k)}. \end{aligned}$$

Therefore, considering the derivatives  $\frac{\partial m_f(\theta)}{\partial \theta_k}$ ,  $\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k}$ ,  $\frac{\partial \text{Im}(\phi_X(t))}{\partial \theta_k}$  and  $\frac{\partial \text{Re}(\phi_X(t))}{\partial \theta_k}$  given in Tables I and III, we can obtain GT moment and trigonometric estimators.

**Normal Inverse Gaussian (NIG) distribution** (see Rachev and Mitnik (2000) and the references therein) A NIG distribution depends on three parameters  $\alpha > 0; \delta > 0; \beta \in (-\alpha, \alpha)$  and its characteristic function is given by:

$$\phi_{NIG}(u; \alpha, \beta, \delta) = \exp \left\{ -\delta \left( \sqrt{\alpha^2 - (\beta + iu)^2} - \sqrt{\alpha^2 - \beta^2} \right) \right\}$$

Thus,  $\text{Re}(\phi_X(u)) = c \cos(d)$ ;  $\text{Im}(\phi_X(u)) = c \sin(d)$ , where

$$c = \exp \left( \delta \sqrt{\alpha^2 - \beta^2} - \delta \left( (\alpha^2 - \beta^2 + u^2)^2 + 4u^2\beta^2 \right)^{0.25} \cos(e) \right),$$

$$d = -\delta \left( (\alpha^2 - \beta^2 + u^2)^2 + 4u^2\beta^2 \right)^{0.25} \sin(e),$$

$$e = 0.5 \arctan \frac{-2u\beta}{\alpha^2 - \beta^2 + u^2}.$$

Moreover, if we assume  $f(X) = X$ , the mean, the variance, the skewness and the kurtosis of Normal Inverse Gaussian distributions are given by:

$$\begin{aligned} m_f(\alpha, \beta, \delta) &= \delta\beta(\alpha^2 - \beta^2)^{-1/2}; & \sigma_f^2 &= \alpha^2\delta(\alpha^2 - \beta^2)^{-3/2}; \\ s_f(\alpha, \beta, \delta) &= \frac{3\beta}{\alpha\sqrt{\delta} \cdot \sqrt[4]{\alpha^2 - \beta^2}}; & k_f(\alpha, \beta, \delta) &= 3 \left( 1 + \frac{\alpha^2 + 4\beta^2}{\delta\alpha^2\sqrt{\alpha^2 - \beta^2}} \right). \end{aligned}$$

The derivatives  $\frac{\partial m_f(\theta)}{\partial \theta_k}$ ,  $\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k}$ ,  $\frac{\partial \text{Im}(\phi_X(t))}{\partial \theta_k}$  and  $\frac{\partial \text{Re}(\phi_X(t))}{\partial \theta_k}$  are given in Tables I and III.

**The Carr, Geman, Madan, Yor (CGMY) distribution** (see Carr *et al.* (2002)) A CGMY distribution depends on four parameters  $C, G, M > 0; Y < 2$  and its characteristic function is given by:

$$\phi_{CGMY}(u; C, G, M, Y) = \exp \left\{ CT(-Y) [(M - iu)^Y - M^Y + (G + iu)^Y - G^Y] \right\}.$$

Thus,  $\text{Re}(\phi_X(t)) = c \cos(d)$ ;  $\text{Im}(\phi_X(t)) = c \sin(d)$ , where

$$\begin{aligned} c &= \exp \left\{ CT(-Y) \left[ -M^Y - G^Y + (M^2 + t^2)^{Y/2} \cos \left( Y \arctan \frac{-t}{M} \right) + \right. \right. \\ &\quad \left. \left. + (G^2 + t^2)^{Y/2} \cos \left( Y \arctan \frac{t}{G} \right) \right] \right\} \end{aligned}$$

**Table III**

*This table summarizes the derivatives used in the trigonometric estimators for NIG, and Stable distributions.*

$S_{\alpha}(\beta, \sigma, \mu)$	where
$\frac{\partial \operatorname{Re}}{\partial \alpha} = -\frac{1}{2} p \left[ \beta (\pi + 2q \ln  t\sigma  + \pi q^2) \operatorname{sgn} t \sin m + 2 \ln  t\sigma  \cos m \right]$	$p =  t\sigma ^{\alpha} e^{-\frac{1}{2} t ^{\alpha}}$
$\frac{\partial \operatorname{Re}}{\partial \beta} = -pq \operatorname{sgn} t \sin m$	$q = \tan \frac{\pi \alpha}{2}$
$\frac{\partial \operatorname{Re}}{\partial \mu} = -te^{-\frac{1}{2} t ^{\alpha}} \sin m$	$m = t\mu + \beta q  t\sigma ^{\alpha} \operatorname{sgn}(t)$
$\frac{\partial \operatorname{Re}}{\partial \sigma} = -\alpha p t \operatorname{sgn}(t\sigma) [\cos m + \beta q \operatorname{sgn} t \sin m]$	
$\frac{\partial \operatorname{Im}}{\partial \alpha} = \frac{1}{2} p \left[ \beta (\pi + 2q \ln  t\sigma  + \pi q^2) \operatorname{sgn} t \cos m - 2 \ln  t\sigma  \sin m \right]$	
$\frac{\partial \operatorname{Im}}{\partial \beta} = pq \operatorname{sgn} t \cos m$	
$\frac{\partial \operatorname{Im}}{\partial \mu} = te^{-\frac{1}{2} t ^{\alpha}} \cos m$	
$\frac{\partial \operatorname{Im}}{\partial \sigma} = -\alpha p t \operatorname{sgn}(t\sigma) [\sin m - \beta q \operatorname{sgn} t \cos m]$	
$NIG(\alpha, \beta, \delta)$	where
$\frac{\partial \operatorname{Re}}{\partial \alpha} = \frac{\alpha \delta w}{nm^3} \left[ n(\beta^2 - \alpha^2 - t^2) \cos q + m^3 \cos(\delta m \sin p) - 2m\beta \sin q \right]$	$m = \sqrt{t^4 + \alpha^4 + \beta^4 + 2(t^2\alpha^2 + t^2\beta^2 - \alpha^2\beta^2)}$
$\frac{\partial \operatorname{Re}}{\partial \beta} = \frac{\delta w}{nm^3} \left[ n\beta(\alpha^2 - \beta^2 - t^2) \cos q - \beta m^3 \cos(\delta m \sin p) + nt(\alpha^2 + \beta^2 + t^2) \sin q \right]$	$n = \sqrt{\alpha^2 - \beta^2}$
$\frac{\partial \operatorname{Re}}{\partial \delta} = w [n \cos(\delta m \sin p) - m \cos q]$	$p = \frac{1}{2} \operatorname{arctg} \frac{2t\beta}{t^2 + \alpha^2 - \beta^2}$
$\frac{\partial \operatorname{Im}}{\partial \alpha} = \frac{\alpha \delta w}{nm^3} \left[ n(\alpha^2 - \beta^2 + t^2) \sin q + m^3 \sin(\delta m \sin p) - 2m\beta \cos q \right]$	$w = e^{\delta(n - m \cos p)}$
$\frac{\partial \operatorname{Im}}{\partial \beta} = \frac{\delta w}{nm^3} \left[ n\beta(\beta^2 - \alpha^2 + t^2) \sin q - \beta m^3 \sin(\delta m \sin p) + nt(\alpha^2 + \beta^2 + t^2) \cos q \right]$	$q = p - \delta m \sin p$
$\frac{\partial \operatorname{Im}}{\partial \delta} = w [n \sin(\delta m \sin p) + m \sin q]$	

$$d = CT(-Y) (M^2 + t^2)^{Y/2} \sin \left( Y \arctan \frac{-t}{M} \right) + \\ + CT(-Y) (G^2 + t^2)^{Y/2} \sin \left( Y \arctan \frac{t}{G} \right).$$

Moreover, if we assume  $f(X) = X$ , the mean, the variance, the skewness and the kurtosis of CGMY distributions are given by:

$$m_f(C, G, M, Y) = C (M^{Y-1} - G^{Y-1}) \Gamma(1 - Y); \\ \sigma_f^2(C, G, M, Y) = C (M^{Y-2} + G^{Y-2}) \Gamma(2 - Y);$$

$$s_f(C, G, M, Y) = \frac{C (M^{Y-3} - G^{Y-3}) \Gamma(3-Y)}{(C (M^{Y-2} + G^{Y-2}) \Gamma(2-Y))^{3/2}};$$

$$k_f(C, G, M, Y) = 3 + \frac{C (M^{Y-4} + G^{Y-4}) \Gamma(4-Y)}{(C (M^{Y-2} + G^{Y-2}) \Gamma(2-Y))^2}.$$

The derivatives  $\frac{\partial m_f(\theta)}{\partial \theta_k}$ ,  $\frac{\partial \sigma_f^2(\theta)}{\partial \theta_k}$ ,  $\frac{\partial \text{Im}(\phi_X(t))}{\partial \theta_k}$  and  $\frac{\partial \text{Re}(\phi_X(t))}{\partial \theta_k}$  are given in Tables I and II.

### 3 An empirical comparison between the GT moment estimator and the maximum likelihood estimator of stable Paretian distributions

In this section we compare the GT moment estimator and the MLE obtained by inverting the characteristic function of stable Paretian distributions (see Rachev and Mittnik (2000)). In particular, we test the above semi-parametric estimator for stable distributions using simulated data. Therefore, using the algorithm proposed by Chambers *et al.* (see Chambers *et al.* (1976)), we generate  $N$  ( $N=200, \dots, 5000$  with step 100) stable distributions  $S_\alpha(\sigma, \beta, \mu)$  with parameters  $\alpha = 0.51, 0.76, 1.26, 1.51, 1.76$ ;  $\beta = -1, -0.5, 0.5, 1$ ;  $\sigma = 1$ ;  $\mu = 1$ . Then we estimate the parameters on the simulated data (for each  $N$ ) considering both estimating methods: the GT moment estimator, and MLE valued inverting the characteristic function with the FFT. As starting point for the GT moment and MLE estimators we use the parameters obtained estimating the series of  $N$  elements with the McCulloch quantile method (see, among others, Rachev and Mittnik (2000)). We obtain the results of GT moment minimizing the sum of the asymptotic variances subject to the usual constraints. This computation requires less time than the MLE approximation. Minimizing the determinant of the asymptotic variance matrix we get more robust results, but we need much more computational time to approximate the parameters.

We measure (on average) the absolute value of the percentage of the distance between the parameters of simulated series and the estimated ones for the different  $\alpha, \sigma, \beta$ , and  $\mu$  i.e., we compute the average (varying  $N$ ) of  $|\Delta\theta_{GT}| = \left| \frac{\theta_{GT} - \theta_{simulation}}{\theta_{simulation}} \right|$ , and similarly of  $|\Delta\theta_{MLE}| = \left| \frac{\theta_{MLE} - \theta_{simulation}}{\theta_{simulation}} \right|$ , where  $\theta = \alpha, \sigma, \beta, \mu$ . These results are given in Table IV where we remark in bold the best approximations. We observe that the sum of the all absolute errors is higher for the MLE. In addition, the empirical analysis shows that the GT moment estimators present a very good performance even in comparison to those obtained with the MLE method.

### 4 Concluding remarks

In the paper we discussed the application of the estimating function method to value the parameters of distributions defined only by their characteristic func-

**Table IV**

This table summarizes the average of the absolute errors we have using either a GT moment estimator or a MLE estimator for different values of  $\alpha$  ( $\alpha = 0.51; 0.76; 1.26; 1.51; 1.76$ ) and  $\beta$  ( $\beta = -1; -0.5; 0.5; 1$ ) and  $\sigma = 1$ ;  $\mu = 1$ .

**Moment**

Alpha	Beta -1				-0.5				0.5				1			
	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $
0.51	0.07	0.28	0.10	0.52	0.04	0.09	0.07	0.12	0.03	0.11	0.09	0.10	0.04	0.09	0.01	0.28
0.76	0.04	0.06	0.01	0.58	0.04	0.07	0.08	0.24	0.04	0.08	0.08	0.21	0.04	0.05	0.01	0.52
1.26	0.04	0.04	0.04	0.43	0.03	0.03	0.09	0.19	0.03	0.03	0.08	0.20	0.04	0.03	0.03	0.40
1.51	0.03	0.03	0.05	0.12	0.03	0.02	0.15	0.07	0.03	0.02	0.14	0.06	0.04	0.03	0.05	0.13
1.76	0.03	0.02	0.09	0.09	0.03	0.03	0.36	0.04	0.03	0.02	0.31	0.04	0.05	0.05	0.07	0.14

**MLE**

Alpha	Beta -1				-0.5				0.5				1			
	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $	$ \Delta\alpha $	$ \Delta\sigma $	$ \Delta\beta $	$ \Delta\mu $
0.51	0.02	0.20	0.62	0.47	0.03	0.25	0.79	0.25	0.03	0.21	0.09	0.05	0.02	0.15	0.02	0.04
0.76	0.23	0.31	0.19	1.54	0.13	0.10	0.50	0.78	0.16	0.13	0.14	0.57	0.23	0.31	0.19	1.54
1.26	0.02	0.02	0.07	0.33	0.02	0.02	0.16	0.24	0.02	0.02	0.08	0.16	0.03	0.02	0.01	0.28
1.51	0.02	0.02	0.06	0.09	0.02	0.02	0.12	0.06	0.02	0.02	0.11	0.05	0.02	0.01	0.00	0.06
1.76	0.02	0.01	0.05	0.04	0.02	0.02	0.18	0.04	0.02	0.02	0.16	0.04	0.02	0.01	0.00	0.04

tions. The proposed methodology showed good versatility, since it could be applied to any bounded function of the underlying random variable. In particular, we propose two EF estimators for the parameters of stable Paretian distributions and other infinitely divisible distributions. Finally we have proposed an empirical comparison based on simulated data of stable Paretian distributions. The good results obtained with the EF moment estimator even with respect to the MLE method, suggest that probably we could make further improvements in parameter estimation using others bounded functions.

**Acknowledgement:** The authors thank for helpful comments seminar audiences at AMAT 2008 (Memphis, USA). For their helpful support in writing the tables, we thank Valentina Acerbis, Marco Cassader and Sebastiano Vitali. Rachev's research was supported by grants from the Division of Mathematical, Life and Physical Science, College of Letters and Science, University of California, Santa Barbara, and the Deutschen Forschungsgemeinschaft. Sergio Ortobelli and Valeria Caviezel were partially supported by grants from ex-murst 60%, 2007 and 2008.

## References

- [1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 1970.

- [2] P. Carr, H. Geman, D.H. Madan and M. Yor, The Fine Structure of Asset Returns, *Journal of Business*, 75, 305-332, (2002).
- [3] J.M. Chambers, C.L. Mallows and B.W. Stuck, A Method for Simulating Stable Random Variables, *Journal of the American Statistical Association*, 71, 340-344, (1976).
- [4] V.P. Godambe, *Estimating Functions*, Oxford University Press, Oxford, 1991.
- [5] V.P. Godambe and M. Thompson, An Extension of Quasi-likelihood Estimation (with discussion), *Journal of Statistical Planning and Inference*, 22, 137-172, (1989).
- [6] Kim Y.S., Rachev S.T., Bianchi M-L., Fabozzi F. Financial Market Models with Levy Processes and Time-Varying Volatility, *Journal of Banking and Finance*, 32/7, 1363-1378, (2008).
- [7] D.X. Li and H.J. Turtle, Semi-parametric ARCH Models: An Estimating Function Approach, *Journal of Business & Economic Statistics*, 18, 174-186, (2000).
- [8] S. Ortobelli and N. Topaloglou, Testing for Preference Orderings Efficiency, *Technical Report*, University of Bergamo, (2008).
- [9] S. Rachev and S. Mitnik, *Stable Paretian Models in Finance*, Wiley & Sons, New York, 2000.
- [10] K. Sato, *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge, 1999.
- [11] M.C.K. Tweedie, An Index which Distinguishes between some Important Exponential Families, in *Statistics: Applications and new Directions*, Proc. Indian statistical institute golden Jubilee International conference (I. Ghosh and J. Roy eds), 1984, 579-604.



# Frequency selective parameterized wavelets of length ten

David W. Roach

Department of Mathematics and Statistics

Murray State University

Murray, KY 42071

david.roach@murraystate.edu

## Abstract

In this paper, the complete parameterization of the length ten wavelets is given with no parameter constraints. Among this class, examples of “frequency selective” wavelets (subband filters) are highlighted that perform comparable to the FBI 9/7 filter as applied to image compression. Specifically, the parameterization is of the dilation equation coefficients for all trigonometric polynomials  $m(\omega)$  which satisfy the necessary conditions for orthogonality, that is  $m(0) = 1$  and  $|m(\omega)|^2 + |m(\omega + \pi)|^2 = 1$ , but with no restriction on the number of vanishing moments except the zeroth which is part of the necessary conditions. Moreover, specific parameters are given that correspond to the Daubechies wavelets, and a frequency response comparison is given showing that there are “flatter/steeper” frequency responses than the Daubechies wavelets but with fewer vanishing moments. We conclude with an image compression scheme to compare the standard wavelets, the FBI 9/7 biorthogonal wavelets with symmetric boundaries, and some “frequency selective” parameterized wavelets.

**Keywords:** Wavelets, coefficients, orthogonal, parameterization

## 1 Introduction

Typically, when someone chooses to use a wavelet in an application, they use the default choices of the standard Daubechies wavelets [3] or the biorthogonal wavelets such as the FBI 9/7 filter [1]. Unfortunately, the wavelet needed for most applications is highly dependent on the data being transformed by the wavelet. In this paper, we give a parameterization for the length ten orthogonal wavelets which allows one to consider a whole continuum of scaling functions that vary from their number of vanishing moments to their regularity. The length of a scaling function is referring directly to the number of nonzero dilation equation coefficients which is one larger than the degree of the associated trigonometric polynomial. The length also refers to the nonzero support of the

scaling function. For instance, the length ten scaling functions have ten dilation coefficients, an associated trigonometric polynomial of degree nine, and is supported on the interval  $[0, 9]$ . The standard Daubechies' wavelets were constructed using a spectral factorization method in such a way that they have minimal length, maximal number of vanishing moments, and minimal phase. For example, the Daubechies' wavelets of lengths two, four, six, eight, and ten with minimal phase each have one, two, three, four, and five vanishing moments respectively. Vanishing moments allow polynomial data within the signal to be well approximated. In [4], Lai and R. constructed explicit parameterizations of all the univariate orthogonal scaling functions of lengths four, six, eight, and ten, but require one to solve a transcendental parameter constraint for both lengths eight and ten. This current paper removes this parameter constraint for the length ten case (refer to [8] for the unconstrained length eight parameterization).

Other researchers have investigated the parameterization of orthogonal wavelets (see [14]). It appears that Schneid and Pittner [12] were the first to give formulas that would lead to the explicit parameterizations for the class of finite length orthogonal scaling functions after finding the Kronecker product of some matrices for wavelet lengths of two through ten. Colella and Heil investigated the length four parameterization in [2]. Others have constructed parameterizations for biorthogonal wavelets as well as multiwavelets (see [11] and [9]). Recently, Regensburger, in [7], constructed the explicit parameterizations for the orthogonal scaling functions with multiple vanishing moments up to length ten by first solving the linear system of equations that result from the vanishing moment conditions and then solving the necessary condition for orthogonality.

This current paper improves on the parameterizations of the past in that the number of vanishing moments is unrestricted and the parameterizations are explicitly stated with no parameter constraints for the class of length ten wavelets.

## 2 Necessary Conditions for Orthogonality

The following necessary conditions for orthogonality are well known in the literature (see [3], [5], and others). Consider a scaling function  $\phi$  that satisfies the dilation equation

$$\phi(x) = \sum_{k=0}^N h_k \phi(2x - k)$$

and its associated trigonometric polynomial  $m$  of degree  $N$  which is given by

$$m(\omega) = \sum_{k=0}^N h_k e^{ik\omega}.$$

It is well known, that  $m$  can be written as an infinite product. In order for this product to converge,  $m$  must not vanish at the origin, i.e.  $m(0) = c \neq 0$ . This

condition immediately implies, where we choose the normalization  $c = 1$ , that

$$\sum_{k=0}^N h_k = 1. \quad (1)$$

Moreover, the necessary condition for the orthogonality of  $\phi$  with its integer shifts is given by

$$|m(\omega)|^2 + |m(\omega + \pi)|^2 = 1. \quad (2)$$

This condition is equivalent to the dilation coefficients satisfying a system of nonlinear equations, specifically

$$\sum_{k=0}^{N-2j} h_k h_{k+2j} = \frac{1}{2} \delta(j), \quad j = 0, \dots, \frac{N-1}{2}$$

where  $\delta(0) = 1$  and  $\delta(j) = 0$  for  $j \neq 0$ .

For the length ten case ( $N = 9$ ) that we are considering currently, we have the following underdetermined nonlinear system:

$$\begin{aligned} h_0^2 + h_1^2 + h_2^2 + h_3^2 + h_4^2 + h_5^2 + h_6^2 + h_7^2 + h_8^2 + h_9^2 &= \frac{1}{2} \\ h_0 h_2 + h_1 h_3 + h_2 h_4 + h_3 h_5 + h_4 h_6 + h_5 h_7 + h_6 h_8 + h_7 h_9 &= 0 \\ h_0 h_4 + h_1 h_5 + h_2 h_6 + h_3 h_7 + h_4 h_8 + h_5 h_9 &= 0 \\ h_0 h_6 + h_1 h_7 + h_2 h_8 + h_3 h_9 &= 0 \\ h_0 h_8 + h_1 h_9 &= 0. \end{aligned}$$

Additionally, these two conditions (1) and (2) imply the zeroth vanishing moment condition  $m(\pi) = 0$  or equivalently the linear equations

$$\sum_{k=0}^{(N-1)/2} h_{2k} = \sum_{k=0}^{(N-1)/2} h_{2k+1} = \frac{1}{2}.$$

Because of the strong patterns between the odd and even indices for the dilation coefficients, it is useful to relabel the dilation equation coefficients in the following fashion

$$m(\omega) = \sum_{k=0}^n a_k e^{2ki\omega} + b_k e^{(2k+1)i\omega}$$

where we let  $n = (N-1)/2$ . Note, since there are no odd length scaling functions satisfying the necessary condition for orthogonality,  $N$  will always be an odd integer.

As a means of summary with our new notation, we conclude with the following statements. Given a scaling function  $\phi$  and its associated trigonometric polynomial

$$m(\omega) = \sum_{k=0}^n a_k e^{2ki\omega} + b_k e^{(2k+1)i\omega}$$

of degree  $2n + 1$ , the necessary condition for orthogonality,

$$|m(\omega)|^2 + |m(\omega + \pi)|^2 = 1,$$

is equivalent to the following system of nonlinear equations:

$$\sum_{k=0}^{n-j} a_k a_{k+j} + b_k b_{k+j} = \frac{1}{2} \delta(j), \quad j = 0, \dots, n-1$$

where  $\delta(0) = 1$  and  $\delta(j) = 0$  for  $j \neq 0$ .

### 3 Length Four

Although the length four parameterization is well known (see [8, 14]), it is used in the construction of the length ten parameterization and is presented here for completeness.

For length four ( $N = 3$  and  $n = 1$ ), the nonlinear system of equations is

$$a_0 + a_1 = \frac{1}{2} \tag{3}$$

$$b_0 + b_1 = \frac{1}{2} \tag{4}$$

$$a_0^2 + a_1^2 + b_0^2 + b_1^2 = \frac{1}{2} \tag{5}$$

$$a_0 a_1 + b_0 b_1 = 0. \tag{6}$$

Subtracting twice equation (6) from equation (5) gives

$$(a_0 - a_1)^2 + (b_0 - b_1)^2 = \frac{1}{2}.$$

This equation allows the introduction of a free parameter, that is

$$a_0 - a_1 = \frac{1}{\sqrt{2}} \sin \theta \tag{7}$$

$$b_0 - b_1 = \frac{1}{\sqrt{2}} \cos \theta. \tag{8}$$

Combining equations (3) and (4) with (7) and (8) gives the length four parameterization

$$\begin{aligned} a_0 &= \frac{1}{4} + \frac{1}{2\sqrt{2}} \sin \theta, & b_0 &= \frac{1}{4} + \frac{1}{2\sqrt{2}} \cos \theta, \\ a_1 &= \frac{1}{4} - \frac{1}{2\sqrt{2}} \sin \theta, & b_1 &= \frac{1}{4} - \frac{1}{2\sqrt{2}} \cos \theta. \end{aligned}$$

These formulas are well known (see [14],[2],[8], and others). To aid in the construction of the longer parameterizations, a different period and phase shift are

chosen for the length four solution, that is  $\theta = 2\alpha - \pi/4$ . With this substitution and some simplification, the length four solution can be written as

$$\begin{aligned} a_0 &= \frac{1}{4}(1 - \cos 2\alpha + \sin 2\alpha) \\ b_0 &= \frac{1}{4}(1 + \cos 2\alpha + \sin 2\alpha) \\ a_1 &= \frac{1}{4}(1 + \cos 2\alpha - \sin 2\alpha) \\ b_1 &= \frac{1}{4}(1 - \cos 2\alpha - \sin 2\alpha) \end{aligned}$$

where this form will simplify future computations. It should be noted that this parameterization is a necessary representation for the coefficients and upon substituting them back into the system of equations (3)-(6), we see that they are also sufficient.

## 4 Length Ten

For the construction of parameterizations for length six and eight see [8]. For length ten ( $N = 9$  and  $n = 4$ ), the nonlinear system of equations is given by

$$a_0 + a_1 + a_2 + a_3 + a_4 = \frac{1}{2} \quad (9)$$

$$b_0 + b_1 + b_2 + b_3 + b_4 = \frac{1}{2} \quad (10)$$

$$a_0^2 + a_1^2 + a_2^2 + a_3^2 + a_4^2 + b_0^2 + b_1^2 + b_2^2 + b_3^2 + b_4^2 = \frac{1}{2} \quad (11)$$

$$a_0a_1 + a_1a_2 + a_2a_3 + a_3a_4 + b_0b_1 + b_1b_2 + b_2b_3 + b_3b_4 = 0 \quad (12)$$

$$a_0a_2 + a_1a_3 + a_2a_4 + b_0b_2 + b_1b_3 + b_2b_4 = 0 \quad (13)$$

$$a_0a_3 + a_1a_4 + b_0b_3 + b_1b_4 = 0 \quad (14)$$

$$a_0a_4 + b_0b_4 = 0. \quad (15)$$

An important step in the construction is establishing the connection between the sums of the even and odd indexed coefficients back to the length four parameterization. More specifically, the sums  $a_0 + a_2 + a_4$ ,  $a_1 + a_3$ ,  $b_0 + b_2 + b_4$ , and  $b_1 + b_3$  satisfy the system of equations associated with the length four parameterization, i.e.

$$(a_0 + a_2 + a_4) + (a_1 + a_3) = \frac{1}{2}$$

$$(b_0 + b_2 + b_4) + (b_1 + b_3) = \frac{1}{2}$$

$$(a_0 + a_2 + a_4)^2 + (a_1 + a_3)^2 + (b_0 + b_2 + b_4)^2 + (b_1 + b_3)^2 = \frac{1}{2}$$

$$(a_0 + a_2 + a_4)(a_1 + a_3) + (b_0 + b_2 + b_4)(b_1 + b_3) = 0.$$

The third equation is equivalent to the sum of equations (11) and (13), and the last one is equivalent to the sum of equations (12) and (14). Therefore, we can use the length four parameterization for these sums, i.e.

$$\begin{aligned} a_0 + a_2 + a_4 &= \frac{1}{4}(1 - \cos 2\alpha + \sin 2\alpha) \\ b_0 + b_2 + b_4 &= \frac{1}{4}(1 + \cos 2\alpha + \sin 2\alpha) \\ a_1 + a_3 &= \frac{1}{4}(1 + \cos 2\alpha - \sin 2\alpha) \\ b_1 + b_3 &= \frac{1}{4}(1 - \cos 2\alpha - \sin 2\alpha). \end{aligned}$$

In an effort to linearize the system of equations, note that the sum and difference of equation (11) and twice equation (15) give the two equations:

$$(a_0 + a_4)^2 + (b_0 + b_4)^2 = \frac{1}{2} - a_1^2 - a_2^2 - a_3^2 - b_1^2 - b_2^2 - b_3^2 \quad (16)$$

$$(a_0 - a_4)^2 + (b_0 - b_4)^2 = \frac{1}{2} - a_1^2 - a_2^2 - a_3^2 - b_1^2 - b_2^2 - b_3^2 := p^2. \quad (17)$$

Although the right hand side,  $p^2$ , has not yet been determined, we use the fact that the right-hand sides of equations (16) and (17) are equivalent and introduce two new free parameters  $\beta$  and  $\gamma$  in the following fashion:

$$\begin{aligned} a_0 + a_4 &= p \cos \beta \\ b_0 + b_4 &= p \sin \beta \\ a_0 - a_4 &= p \cos \gamma \\ b_0 - b_4 &= p \sin \gamma. \end{aligned}$$

There are now 8 linear equations and ten unknowns. For the last two equations, we introduce two additional free parameters  $q$  and  $r$  for the differences between the odd indices and rotate them using an orthogonal matrix which depends on the parameter  $\gamma$ , i.e.,

$$\begin{aligned} a_1 - a_3 &= q \cos \gamma - r \sin \gamma \\ b_1 - b_3 &= q \sin \gamma + r \cos \gamma. \end{aligned}$$

The decision to rotate these free parameters by the orthogonal matrix depending on  $\gamma$  was based on the nonlinear relationship in equation (14) which involves  $\cos \gamma$  and  $\sin \gamma$ .

Now, solving the linear system of equations involving the free parameters  $\alpha, \beta, \gamma, p, q$ , and  $r$  which necessarily satisfy the nonlinear system of equations yields

$$\begin{aligned}
a_0 &= \frac{p}{2}(\cos \beta + \cos \gamma) \\
b_0 &= \frac{p}{2}(\sin \beta + \sin \gamma) \\
a_1 &= \frac{1}{8}(1 + \cos 2\alpha - \sin 2\alpha) + \frac{1}{2}(q \cos \gamma - r \sin \gamma) \\
b_1 &= \frac{1}{8}(1 - \cos 2\alpha - \sin 2\alpha) + \frac{1}{2}(q \sin \gamma + r \cos \gamma) \\
a_2 &= \frac{1}{4}(1 - \cos 2\alpha + \sin 2\alpha) - p \cos \beta \\
b_2 &= \frac{1}{4}(1 + \cos 2\alpha + \sin 2\alpha) - p \sin \beta \\
a_3 &= \frac{1}{8}(1 + \cos 2\alpha - \sin 2\alpha) - \frac{1}{2}(q \cos \gamma - r \sin \gamma) \\
b_3 &= \frac{1}{8}(1 - \cos 2\alpha - \sin 2\alpha) - \frac{1}{2}(q \sin \gamma + r \cos \gamma) \\
a_4 &= \frac{p}{2}(\cos \beta - \cos \gamma) \\
b_4 &= \frac{p}{2}(\sin \beta - \sin \gamma).
\end{aligned}$$

As far as the nonlinear equations, notice that equation (15) is satisfied immediately. Now plugging the linear solutions into equation (14), we get

$$\begin{aligned}
a_0 a_3 + a_1 a_4 + b_0 b_3 + b_1 b_4 &= 0 \\
\frac{p}{8}(-4q + \cos \beta + \sin \beta + \cos(2\alpha + \beta) - \sin(2\alpha + \beta)) &= 0
\end{aligned}$$

and since  $p = 0$  would result in a constrained solution, we have the parameterization of  $q$  as

$$\begin{aligned}
q &= \frac{1}{4}(\cos \beta + \sin \beta + \cos(2\alpha + \beta) - \sin(2\alpha + \beta)) \\
&= \frac{1}{2}(\cos \alpha - \sin \alpha) \cos(\alpha + \beta).
\end{aligned}$$

Using this value of  $q$  and substituting the linear equation solutions into equation (12), we see that equation (12) is now satisfied while equation (13) produces a quadratic equation in terms of the unknown  $p$ , that is

$$\begin{aligned}
a_0 a_2 + a_1 a_3 + a_2 a_4 + b_0 b_2 + b_1 b_3 + b_2 b_4 &= 0 \\
p^2 - \frac{1}{2}(\cos \alpha + \sin \alpha) \sin(\alpha + \beta) p + \dots & \\
\frac{1}{64}(16r^2 + 2 \cos(2(\alpha + \beta)) - \sin(2(\alpha + \beta)) + 2 \sin 2\alpha + \sin 2\beta - 2) &= 0
\end{aligned}$$

which has a solution in terms of the free parameter  $r$  of

$$p = \frac{1}{4} \left( (\cos \alpha + \sin \alpha) \sin(\alpha + \beta) \pm \sqrt{1 - 4r^2 - \cos(2(\alpha + \beta))} \right)$$

with a constraint on the size of the parameter  $r$  of

$$r^2 \leq \frac{1 - \cos(2(\alpha + \beta))}{4} \quad (18)$$

$$= \frac{1}{2} \sin^2(\alpha + \beta) \quad (19)$$

that is necessary to keep the parameter  $p$  real. So, we introduce the free parameter  $\delta$  for  $r$  that satisfies the inequality (18), i.e.

$$r = \frac{1}{\sqrt{2}} \cos \delta \sin(\alpha + \beta).$$

After substituting this parameterization of  $r$  back into the expression for  $p$ , we have

$$p = \frac{1}{4} \sin(\alpha + \beta) (\cos \alpha + \sin \alpha - \sqrt{2} \sin \delta).$$

Substituting these parameterizations of  $p, q$ , and  $r$  back into the nonlinear system, shows that in fact these necessary conditions on the parameters are also sufficient. Therefore the complete parameterization of all coefficient vectors of length ten that satisfy the necessary conditions for orthogonality can be parameterized by the following:

$$\begin{aligned} p &= \frac{1}{4} \sin(\alpha + \beta) (\cos \alpha + \sin \alpha - \sqrt{2} \sin \delta) \\ q &= \frac{1}{2} (\cos \alpha - \sin \alpha) \cos(\alpha + \beta) \\ r &= \frac{1}{\sqrt{2}} \cos \delta \sin(\alpha + \beta) \\ a_0 &= \frac{p}{2} (\cos \beta + \cos \gamma) \\ b_0 &= \frac{p}{2} (\sin \beta + \sin \gamma) \\ a_1 &= \frac{1}{8} (1 + \cos 2\alpha - \sin 2\alpha) + \frac{1}{2} (q \cos \gamma - r \sin \gamma) \\ b_1 &= \frac{1}{8} (1 - \cos 2\alpha - \sin 2\alpha) + \frac{1}{2} (q \sin \gamma + r \cos \gamma) \\ a_2 &= \frac{1}{4} (1 - \cos 2\alpha + \sin 2\alpha) - p \cos \beta \\ b_2 &= \frac{1}{4} (1 + \cos 2\alpha + \sin 2\alpha) - p \sin \beta \\ a_3 &= \frac{1}{8} (1 + \cos 2\alpha - \sin 2\alpha) - \frac{1}{2} (q \cos \gamma - r \sin \gamma) \\ b_3 &= \frac{1}{8} (1 - \cos 2\alpha - \sin 2\alpha) - \frac{1}{2} (q \sin \gamma + r \cos \gamma) \\ a_4 &= \frac{p}{2} (\cos \beta - \cos \gamma) \\ b_4 &= \frac{p}{2} (\sin \beta - \sin \gamma) \end{aligned}$$



Wavelet	$\alpha$	$\beta$	$\gamma$	$\delta$
Haar	-2.35619449	-2.35619449	-2.35619449	1.57079633
D4	-2.61799388	-2.09439510	-2.09439510	1.30899694
D6	-2.85603576	-1.96184952	-1.96184952	1.07095506
D8	-3.08353313	-1.88256239	-1.88256239	0.93539307
D10	-3.30496928	-1.80914574	-1.85080961	0.88683222
OJ10	-3.27396878	-1.82536723	-1.87718786	0.93186197
T10	0.71389116	-2.86392827	1.66872168	2.43807534
MF10	0.74143253	-2.94548381	1.51834616	2.20225600

Table 1: Parameters associated with the standard Daubechies wavelets and some other length ten wavelets.

which is well-defined for any parameters  $\alpha, \beta, \gamma$ , and  $\delta \in \mathbf{IR}$ . This parameterization contains all of the standard Daubechies wavelets of length ten or less where the parameters needed for each are given in Table 1 along with some other wavelets which will be used in the image compression comparison. Figure 1 shows the scaling functions associated with the parameters chosen in Table 1.

## 5 Frequency Selective Wavelets and Image Compression

In Daubechies seminal work [3], an emphasis is placed on discrete orthogonal wavelets which have a maximum number of vanishing moments. In that work, an example is put forth where one of the vanishing moments is moved away from  $\omega = -\pi$  towards  $-\pi/2$  and the comment is made that this wavelet is much closer to a “realistic” subband coding filter (see pg. 201 in [3]). In a later work, Ojanen in [6] develops a scheme to find the maximal regularity for the scaling functions based on moving a finite number vanishing moments towards the center. Ojanen lists one optimal length 10 wavelet which we will call OJ10(see Table 1 for the necessary parameters). The frequency response graphs in Figure 2 and 3 illustrate how the filters will treat various frequencies in the range  $[-\pi, \pi]$ . An “ideal” filter (which can not be implemented with a finite convolution) has a frequency response that is one on the interval  $[-\pi/2, \pi/2]$  and zero otherwise. In Figure 3b, notice that OJ10 uses a zero near -2.6 to improve the “flatness” in this area and subsequently giving it a steeper transition. In the literature, band-pass filters that model the “ideal” filter are called “frequency selective”. T10 and MF10 both have only the zeroth vanishing moment, but MF10 keeps a shallow profile in the first part of the interval and achieves a steeper transition than D10 or OJ10. T10 is highlighted here because of its extreme steepness at the transition compared to the others, but sacrifices flatness in the first part of the interval.

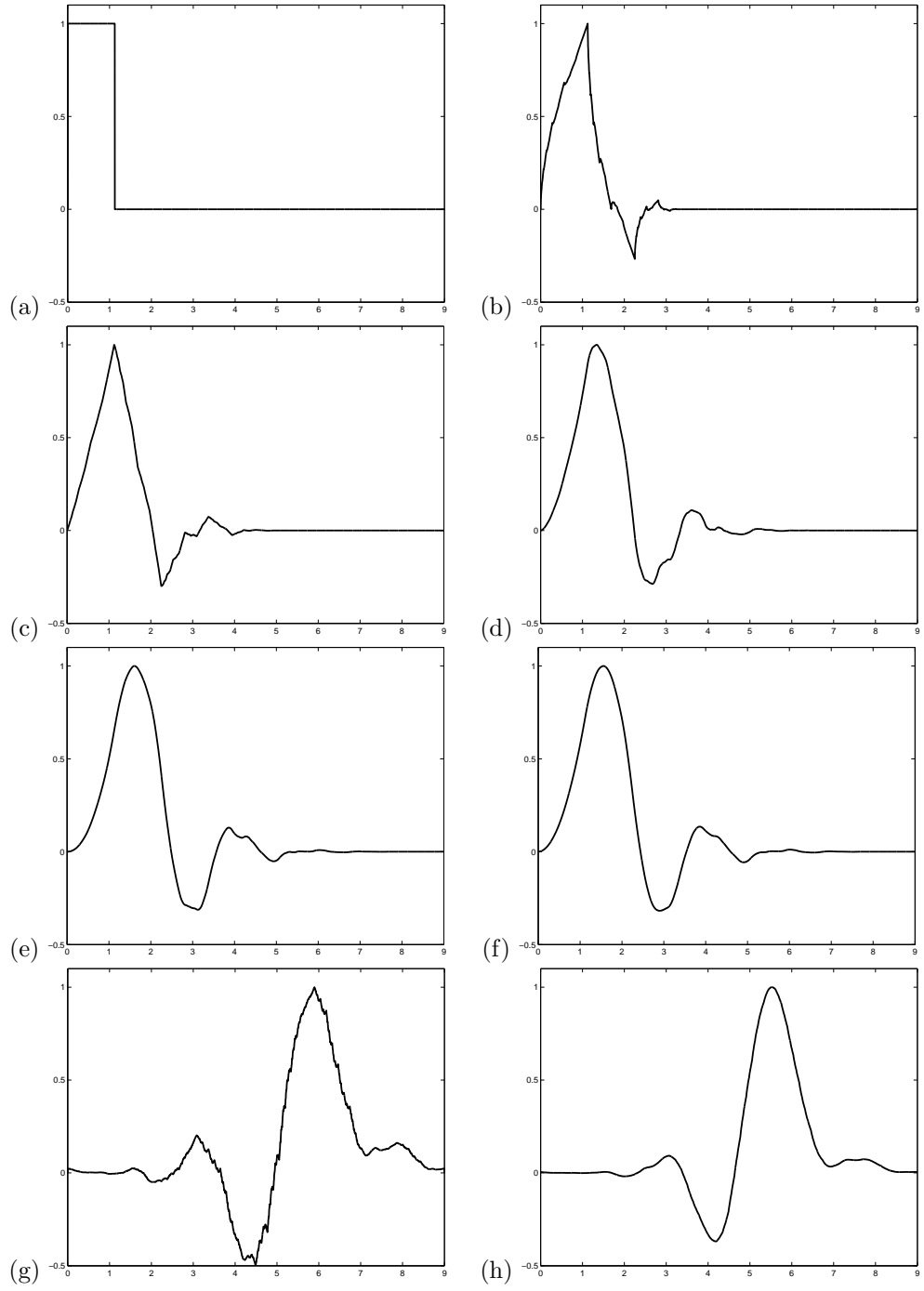


Figure 1: Scaling function plots for the specific parameters in Table 1. (a) Haar, (b) D4, (c) D6, (d) D8, (e) D10, (f) OJ10, (g) T10, and (h) MF10.

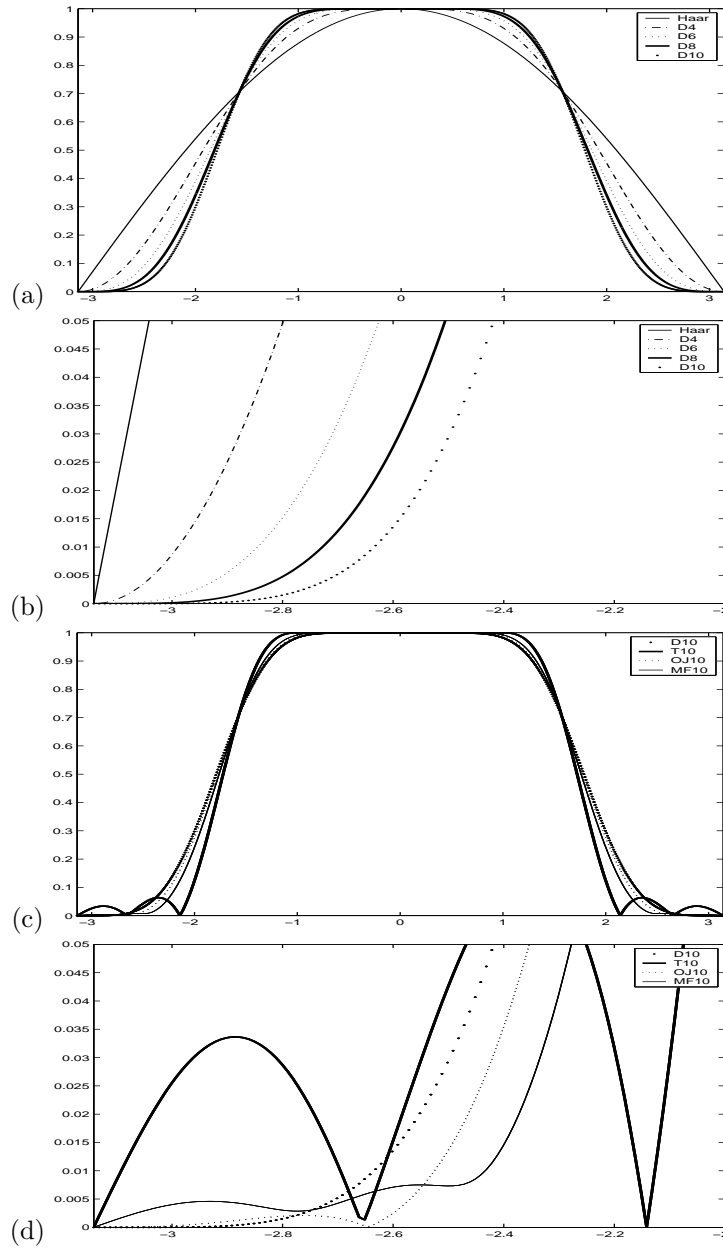


Figure 2: Frequency response plots  $|m(\omega)|$  for the scaling functions from Table 1. (a) Haar, D4, D6, D8, and, D10 (b) zoomed in plot of previous, (c) D10, T10, OJ10, and MF10, (d) zoomed in plot of previous.

Row=469							
Wavelet	Barb	Boat	Lena	Marm	Bark	Fing	Sand
D10	3034	1052	820	73	2973	1676	1943
OJ10	3039	1050	813	73	2946	1675	1954
T10	2822	<b>957</b>	880	95	<b>2610</b>	1670	1993
MF10	<b>2793</b>	960	846	82	2640	1669	1959
FBI 9/7	2900	976	<b>798</b>	<b>58</b>	2795	<b>1668</b>	<b>1825</b>
Row=191							
Wavelet	Barb	Boat	Lena	Marm	Bark	Fing	Sand
D10	2064	1651	951	113	2384	1544	2317
OJ10	<b>2029</b>	1665	943	<b>105</b>	2358	1532	2311
T10	2409	1603	986	256	2210	<b>1529</b>	2255
MF10	2379	1603	948	163	<b>2188</b>	1532	2241
FBI 9/7	2166	<b>1593</b>	<b>880</b>	106	2198	1548	<b>2090</b>
Row=287							
Wavelet	Barb	Boat	Lena	Marm	Bark	Fing	Sand
D10	2818	1799	929	76	2407	1629	1972
OJ10	<b>2799</b>	1785	917	<b>70</b>	2398	1616	1960
T10	3163	1829	946	184	2272	<b>1555</b>	2126
MF10	3172	1826	941	116	2259	1568	2099
FBI 9/7	2856	<b>1756</b>	<b>854</b>	90	<b>2182</b>	1634	<b>1911</b>

Table 2: The  $\ell_1$  norm of the wavelet coefficients for a one-level decomposition of a single row from some test images. The lowest  $\ell_1$  norm has been boldfaced for convenience.

In our first numerical experiment, we compare the four length-ten filters D10, OJ10, T10, and MF10 with the biorthogonal FBI 9/7 wavelet with a single level decomposition of a row from our test images and compute the  $\ell_1$  norm of the wavelet coefficients. A lower  $\ell_1$  norm would suggest a more efficient representation for the vector. We chose three random rows and tested all five wavelets on seven different images using the specified row. As can be seen in Table 2, the 9/7 filter appears to do better more than half of the time. The lowest  $\ell_1$  norm has been boldfaced for each image. The results are not conclusive, but are still interesting. In a second more comprehensive numerical experiment, we present an image compression comparison for the wavelets D10, OJ10, T10, MF10, and the FBI biorthogonal 9/7 wavelet with symmetric boundary extensions. Because of its common use as an industry standard and similar length, we chose to include the biorthogonal 9/7 wavelet in the comparison. The 9/7 biorthogonal wavelet has one advantage over the the length ten parameterization in that it is symmetric. This symmetry can be used to improve the performance of image compression at the image boundaries. We have included this advantage in our numerical results. The details of the numerical experiment are as follows:

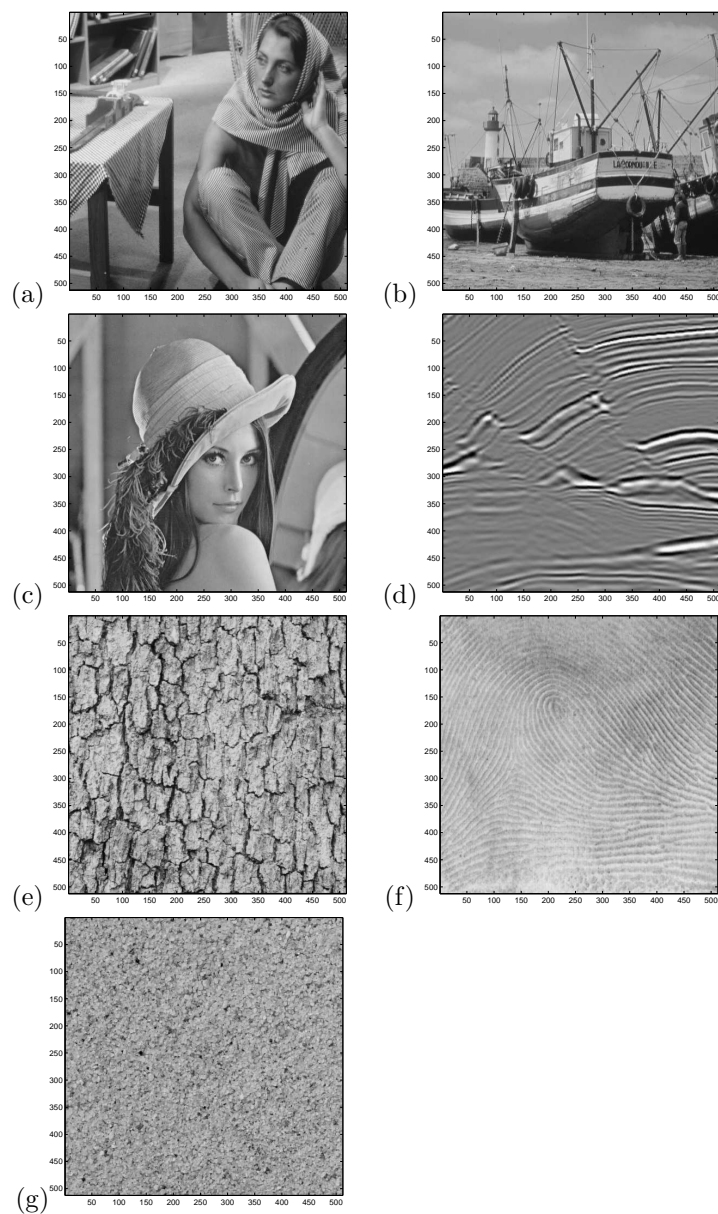


Figure 3: The seven test images used in the compression scheme ( $512 \times 512$  grayscale images): (a) Barb, (b) Boat, (c) Lena, (d) Marm, (e) Bark, (f) Fing, and (g) Sand.

Wavelet	Barb	Boat	Lena	Marm	Bark	Fing	Sand
D10	26.29	28.74	32.30	34.78	21.27	30.14	23.27
OJ10	26.34	28.76	32.32	34.93	21.31	30.18	23.28
T10	<b>26.43</b>	28.59	32.11	35.01	21.41	<b>30.35</b>	23.42
MF10	<b>26.43</b>	28.69	32.43	<b>35.57</b>	<b>21.44</b>	30.28	23.44
FBI9/7	26.30	<b>29.22</b>	<b>32.81</b>	35.11	21.43	30.14	<b>23.46</b>

Table 3: PSNR results for the seven images Barb, Boat, Lena, Marm, Bark, Fing, and Sand using the wavelets D10, OJ10, T10, MF10, and the FBI biorthogonal 9/7. The best PSNR for each image is boldfaced.

- Seven level decomposition with periodic boundaries (except for 9/7 which has symmetric extensions) using D10, OJ10, T10, MF10, and 9/7 FBI. Note: All of the scaling functions presented in this paper need to be normalized by a multiple of  $\sqrt{2}$  during implementation.
- Embedded Zero-tree (EZW) compression (see [13] and [10]) with a file size ratio of 32:1. For this experiment, all of the images are  $512 \times 512$  with a PGM file-size of 256Kb and a compressed file-size of 8Kb. This particular EZW implementation is not completely optimized and would not necessarily yield the maximum PSNR possible but serves well as a comparative measure of the true compressibility of the wavelet decomposition.
- Seven level reconstruction followed by a Peak Signal to Noise Ratio (PSNR), i.e.

$$RMSE = \sqrt{\frac{1}{512^2} \sum_{i=1}^{512} \sum_{j=1}^{512} |A_{i,j} - \tilde{A}_{i,j}|^2}$$

$$PSNR = 20 \log_{10} \left( \frac{255}{RMSE} \right)$$

where  $A_{i,j}$  is the original matrix of grayscale values and  $\tilde{A}_{i,j}$  is the compressed version.

The results from the experiment are given in Table 3.

## References

- [1] Brislawn, C., J. Bradley, R. Onyshczak, and T. Hopper, The FBI compression standard for digitized fingerprint images, Appl. Digital Image Process., Proc. SPIE **2847** (1996), 344-355.
- [2] D. Colella and C. Heil, The characterization of continuous, four-coefficient scaling functions and wavelets, IEEE Trans. Inf. Th., Special Issue on

- Wavelet Transforms and Multiresolution Signal Analysis, 38 (1992), pp. 876-881.
- [3] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.
  - [4] M. J. Lai and D. W. Roach, Parameterizations of univariate orthogonal wavelets with short support, *Approximation Theory X: Splines, Wavelets, and Applications*, C. K. Chui, L. L. Schumaker, and J. Stöckler (eds.), Vanderbilt University Press, Nashville, 2002, 369-384.
  - [5] W. Lawton, Necessary and sufficient conditions for constructing orthonormal wavelet bases, *J. Math. Phys.* **32** (1991), 57-61.
  - [6] H. Ojanen, Orthonormal compactly supported wavelets with optimal sobolev regularity, *Applied and Computational Harmonic Analysis* 10, pp. 93-98 (2001).
  - [7] G. Regensburger, Parametrizing compactly supported orthonormal wavelets by discrete moments. *Appl. Algebra Eng., Commun. Comput.* 18, 6 (Nov. 2007), 583-601.
  - [8] D. W. Roach, The Parameterization of the Length Eight Orthogonal Wavelets with No Parameter Constraints, *Approximation Theory XII: San Antonio 2007*, M. Neamtu and L. Schumaker (eds.), Nashboro Press, pp. 332-347, 2008.
  - [9] H. L. Resnikoff, J. Tian, R. O. Wells, Jr., Biorthogonal wavelet space: parametrization and factorization, *SIAM J. Math. Anal.* **33** (2001), no. 1, 194-215.
  - [10] A. Said and W. A. Pearlman, A new fast and efficient image codec based on set partitioning in hierarchical trees, *IEEE Transactions on Circuits and Systems for Video Technology* **6** (1996), 243-250.
  - [11] J. Qingtang, Paramterization of m-channel orthogonal multifilter banks, *Advances in Computational Mathematics* **12** (2000), 189-211.
  - [12] J. Schneid and S. Pittner, On the parametrization of the coefficients of dilation equations for compactly supported wavelets, *Computing* **51** (1993), 165-173.
  - [13] J. M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, *IEEE Transactions Signal Processing* **41** (1993), 3445-3462.
  - [14] R. O. Wells, Jr., Parameterizing smooth compactly supported wavelets, *Trans. Amer. Math. Soc.* **338** (1993), 919-931.
  - [15] M. V. Wickerhauser, Comparison of picture compression methods: wavelet, wavelet packet, and local cosine transform coding, *Wavelets: theory, algorithms, and applications (Taormina, 1993)*, Academic Press, San Diego, CA, 1994, 585-621.

---

**Instructions to Contributors**  
**Journal of Concrete and Applicable Mathematics**  
 A quarterly international publication of Eudoxus Press, LLC, of TN.

**Editor in Chief: George Anastassiou**  
 Department of Mathematical Sciences  
 University of Memphis  
 Memphis, TN 38152-3240, U.S.A.

**1. Manuscripts hard copies in triplicate, and in English, should be submitted to the Editor-in-Chief:**

**Prof. George A. Anastassiou**  
 Department of Mathematical Sciences  
 The University of Memphis  
 Memphis, TN 38152, USA.  
 Tel. 901.678.3144  
 e-mail: [ganastss@memphis.edu](mailto:ganastss@memphis.edu)

Authors may want to recommend an associate editor the most related to the submission to possibly handle it.

Also authors may want to submit a list of six possible referees, to be used in case we cannot find related referees by ourselves.

**2. Manuscripts should be typed using any of TEX, LaTeX, AMS-TEX, or AMS-LaTeX and according to EUDOXUS PRESS, LLC. LATEX STYLE FILE. (Click [HERE](#) to save a copy of the style file.) They should be carefully prepared in all respects. Submitted copies should be brightly printed (not dot-matrix), double spaced, in ten point type size, on one side high quality paper 8(1/2)x11 inch. Manuscripts should have generous margins on all sides and should not exceed 24 pages.**

**3. Submission is a representation that the manuscript has not been published previously in this or any other similar form and is not currently under consideration for publication elsewhere. A statement transferring from the authors (or their employers, if they hold the copyright) to Eudoxus Press, LLC, will be required before the manuscript can be accepted for publication. The Editor-in-Chief will supply the necessary forms for this transfer. Such a written transfer of copyright, which previously was assumed to be implicit in the act of submitting a manuscript, is necessary under the U.S. Copyright Law in order for the publisher to carry through the dissemination of research results and reviews as widely and effectively as possible.**



**4. The paper starts with the title of the article, author's name(s) (no titles or degrees), author's affiliation(s) and e-mail addresses. The affiliation should comprise the department, institution (usually university or company), city, state (and/or nation) and mail code.**

**The following items, 5 and 6, should be on page no. 1 of the paper.**

**5. An abstract is to be provided, preferably no longer than 150 words.**

**6. A list of 5 key words is to be provided directly below the abstract. Key words should express the precise content of the manuscript, as they are used for indexing purposes.**

**The main body of the paper should begin on page no. 1, if possible.**

**7. All sections should be numbered with Arabic numerals (such as: 1. INTRODUCTION) .**

**Subsections should be identified with section and subsection numbers (such as 6.1. Second-Value Subheading).**

**If applicable, an independent single-number system (one for each category) should be used to label all theorems, lemmas, propositions, corollaries, definitions, remarks, examples, etc. The label (such as Lemma 7) should be typed with paragraph indentation, followed by a period and the lemma itself.**

**8. Mathematical notation must be typeset. Equations should be numbered consecutively with Arabic numerals in parentheses placed flush right, and should be thusly referred to in the text [such as Eqs.(2) and (5)]. The running title must be placed at the top of even numbered pages and the first author's name, et al., must be placed at the top of the odd numbered pages.**

**9. Illustrations (photographs, drawings, diagrams, and charts) are to be numbered in one consecutive series of Arabic numerals. The captions for illustrations should be typed double space. All illustrations, charts, tables, etc., must be embedded in the body of the manuscript in proper, final, print position. In particular, manuscript, source, and PDF file version must be at camera ready stage for publication or they cannot be considered.**

**Tables are to be numbered (with Roman numerals) and referred to by number in the text. Center the title above the table, and type explanatory footnotes (indicated by superscript lowercase letters) below the table.**

**10. List references alphabetically at the end of the paper and number them consecutively. Each must be cited in the text by the appropriate Arabic numeral in square brackets on the baseline.**

**References should include (in the following order):  
initials of first and middle name, last name of author(s)  
title of article,**

name of publication, volume number, inclusive pages, and year of publication.

Authors should follow these examples:

### **Journal Article**

1. H.H.Gonska, Degree of simultaneous approximation of bivariate functions by Gordon operators, (journal name in italics) *J. Approx. Theory*, 62,170-191(1990).

### **Book**

2. G.G.Lorentz, (title of book in italics) *Bernstein Polynomials* (2nd ed.), Chelsea, New York, 1986.

### **Contribution to a Book**

3. M.K.Khan, Approximation properties of beta operators, in (title of book in italics) *Progress in Approximation Theory* (P.Nevai and A.Pinkus, eds.), Academic Press, New York, 1991, pp.483-495.

11. All acknowledgements (including those for a grant and financial support) should occur in one paragraph that directly precedes the References section.

12. Footnotes should be avoided. When their use is absolutely necessary, footnotes should be numbered consecutively using Arabic numerals and should be typed at the bottom of the page to which they refer. Place a line above the footnote, so that it is set off from the text. Use the appropriate superscript numeral for citation in the text.

13. After each revision is made please again submit three hard copies of the revised manuscript, including in the final one. And after a manuscript has been accepted for publication and with all revisions incorporated, manuscripts, including the TEX/LaTeX source file and the PDF file, are to be submitted to the Editor's Office on a personal-computer disk, 3.5 inch size. Label the disk with clearly written identifying information and properly ship, such as:

Your name, title of article, kind of computer used, kind of software and version number, disk format and files names of article, as well as abbreviated journal name.

Package the disk in a disk mailer or protective cardboard. Make sure contents of disks are identical with the ones of final hard copies submitted!

Note: The Editor's Office cannot accept the disk without the accompanying matching hard copies of manuscript. No e-mail final submissions are allowed! The disk submission must be used.

14. Effective 1 Nov. 2009 for current journal page charges, contact the Editor in Chief. Upon acceptance of the paper an invoice will be sent to the contact author. The fee payment will be due one month from the invoice date. The article will proceed to publication only after the fee is paid. The charges are to be sent, by money order or certified check, in US dollars, payable to Eudoxus Press, LLC, to the address shown on

the Eudoxus [homepage](#).

No galleys will be sent and the contact author will receive one(1) electronic copy of the journal issue in which the article appears.

15. This journal will consider for publication only papers that contain proofs for their listed results.

# **TABLE OF CONTENTS, JOURNAL OF CONCRETE AND APPLICABLE MATHEMATICS, VOL. 8, NO. 1, 2010**

<b>Iterative reconstruction and stability bounds for sampling models, Ernesto Acosta-Reyes,.....</b>	<b>9</b>
<b>Global existence and blow up for solutions to higher order Boussinesq systems, De Godefroy Akmel,.....</b>	<b>24</b>
<b>Caputo Fractional Multivariate Opial type inequalities on spherical shells, George A. Anastassiou,.....</b>	<b>41</b>
<b>Asymptotics for Szego polynomials with respect to a class of weakly convergent measures, Michael Arciero, Lewis Pakula,.....</b>	<b>92</b>
<b>A computational approach to the determination of nets, Hans Fetter, Juan H. Arredondo R.,.....</b>	<b>98</b>
<b>Kernel based Wavelets on <math>S^3</math>, S. Bernstein, S. Ebert, .....</b>	<b>110</b>
<b>A Taste of Ideal Projectors, Boris Shekhtman,.....</b>	<b>125</b>
<b>Semiparametric estimators for heavy tailed distributions, Valeria Caviezel et al,</b>	<b>150</b>
<b>Frequency selective parameterized wavelets of length ten, David W. Roach,.....</b>	<b>165</b>

**VOLUME 8, NUMBER 2      APRIL 2010**

**ISSN:1548-5390 PRINT,1559-176X ONLINE**



**JOURNAL  
OF CONCRETE  
AND APPLICABLE**

**MATHEMATICS  
SPECIAL ISSUE II :APPLIED MATHEMATICS  
AND APPROXIMATION THEORY**

**EUDOXUS PRESS,LLC**

**SCOPE AND PRICES OF THE JOURNAL**  
**Journal of Concrete and Applicable Mathematics**

A quartely international publication of **Eudoxus Press,LLC**

**Editor in Chief: George Anastassiou**

Department of Mathematical Sciences,  
 University of Memphis  
 Memphis, TN 38152, U.S.A.  
 ganastss@memphis.edu

The main purpose of the "Journal of Concrete and Applicable Mathematics" is to publish high quality original research articles from all subareas of Non-Pure and/or Applicable Mathematics and its many real life applications, as well connections to other areas of Mathematical Sciences, as long as they are presented in a Concrete way. It welcomes also related research survey articles and book reviews. A sample list of connected mathematical areas with this publication includes and is not restricted to: Applied Analysis, Applied Functional Analysis, Probability theory, Stochastic Processes, Approximation Theory, O.D.E, P.D.E, Wavelet, Neural Networks, Difference Equations, Summability, Fractals, Special Functions, Splines, Asymptotic Analysis, Fractional Analysis, Inequalities, Moment Theory, Numerical Functional Analysis, Tomography, Asymptotic Expansions, Fourier Analysis, Applied Harmonic Analysis, Integral Equations, Signal Analysis, Numerical Analysis, Optimization, Operations Research, Linear Programming, Fuzzyness, Mathematical Finance, Stochastic Analysis, Game Theory, Math. Physics aspects, Applied Real and Complex Analysis, Computational Number Theory, Graph Theory, Combinatorics, Computer Science Math. related topics, combinations of the above, etc. In general any kind of Concretely presented Mathematics which is Applicable fits to the scope of this journal.

Working Concretely and in Applicable Mathematics has become a main trend in many recent years, so we can understand better and deeper and solve the important problems of our real and scientific world.

"Journal of Concrete and Applicable Mathematics" is a peer-reviewed International Quarterly Journal.

We are calling for papers for possible publication. The contributor should send three copies of the contribution to the editor in-Chief typed in TEX, LATEX double spaced. [ See: Instructions to Contributors]

**Journal of Concrete and Applicable Mathematics(JCAAM)**

**ISSN:1548-5390 PRINT, 1559-176X ONLINE.**

is published in January, April, July and October of each year by

**EUDOXUS PRESS,LLC,**

1424 Beaver Trail Drive, Cordova, TN38016, USA,

Tel. 001-901-751-3553

anastassioug@yahoo.com

<http://www.EudoxusPress.com>.

**Visit also [www.msci.memphis.edu/~ganastss/jcaam](http://www.msci.memphis.edu/~ganastss/jcaam).**

**Webmaster: Ray Clapsadle**

**Annual Subscription Current Prices:** For USA and Canada, Institutional: Print \$400, Electronic \$250, Print and Electronic \$450. Individual: Print \$150, Electronic

\$80,Print &Electronic \$200.For any other part of the world add \$50 more to the above prices for Print.

Single article PDF file for individual \$15.Single issue in PDF form for individual \$60.

No credit card payments.Only certified check,money order or international check in US dollars are acceptable.

Combination orders of any two from JoCAAA,JCAAM,JAFa receive 25% discount,all three receive 30% discount.

**Copyright**©2010 by Eudoxus Press,LLC all rights reserved.JCAAM is printed in USA.

**JCAAM is reviewed and abstracted by AMS Mathematical Reviews,MATHSCI,and Zentralblatt MATH.**

It is strictly prohibited the reproduction and transmission of any part of JCAAM and in any form and by any means without the written permission of the publisher.It is only allowed to educators to Xerox articles for educational purposes.The publisher assumes no responsibility for the content of published papers.

***JCAAM IS A JOURNAL OF RAPID PUBLICATION***

---

## Editorial Board

### Associate Editors

---

**Editor in -Chief:**

George Anastassiou  
 Department of Mathematical Sciences  
 The University Of Memphis  
 Memphis, TN 38152, USA  
 tel. 901-678-3144, fax 901-678-2480  
 e-mail ganastss@memphis.edu  
[www.msci.memphis.edu/~anastasg/anlyjour.htm](http://www.msci.memphis.edu/~anastasg/anlyjour.htm)  
 Areas: Approximation Theory,  
 Probability, Moments, Wavelet,  
 Neural Networks, Inequalities, Fuzzyness.

**Associate Editors:**

1) Ravi Agarwal  
 Florida Institute of Technology  
 Applied Mathematics Program  
 150 W. University Blvd.  
 Melbourne, FL 32901, USA  
[agarwal@fit.edu](mailto:agarwal@fit.edu)  
 Differential Equations, Difference  
 Equations,  
 Inequalities

2) Drumi D. Bainov  
 Medical University of Sofia  
 P.O. Box 45, 1504 Sofia, Bulgaria  
[drumibainov@yahoo.com](mailto:drumibainov@yahoo.com)  
 Differential Equations, Optimal Control,  
 Numerical Analysis, Approximation Theory

3) Carlo Bardaro  
 Dipartimento di Matematica & Informatica  
 Università di Perugia  
 Via Vanvitelli 1  
 06123 Perugia, ITALY  
 tel. +390755855034, +390755853822,  
 fax +390755855024  
[bardaro@unipg.it](mailto:bardaro@unipg.it) ,  
[bardaro@dipmat.unipg.it](mailto:bardaro@dipmat.unipg.it)  
 Functional Analysis and Approximation Th.,  
 Summability, Signal Analysis, Integral  
 Equations,  
 Measure Th., Real Analysis

4) Francoise Bastin  
 Institute of Mathematics  
 University of Liege  
 4000 Liege

21) Gustavo Alberto Perla Menzala  
 National Laboratory of Scientific Computation  
 LNCC/MCT  
 Av. Getulio Vargas 333  
 25651-075 Petropolis, RJ  
 Caixa Postal 95113, Brasil  
 and

Federal University of Rio de Janeiro  
 Institute of Mathematics  
 RJ, P.O. Box 68530 Rio de Janeiro, Brasil  
[perla@lncc.br](mailto:perla@lncc.br) and [perla@im.ufrj.br](mailto:perla@im.ufrj.br)  
 Phone 55-24-22336068, 55-21-25627513 Ext 224  
 FAX 55-24-22315595  
 Hyperbolic and Parabolic Partial Differential  
 Equations,  
 Exact controllability, Nonlinear Lattices and  
 Global  
 Attractors, Smart Materials

22) Ram N. Mohapatra  
 Department of Mathematics  
 University of Central Florida  
 Orlando, FL 32816-1364  
 tel. 407-823-5080  
[ramm@pegasus.cc.ucf.edu](mailto:ramm@pegasus.cc.ucf.edu)  
 Real and Complex analysis, Approximation Th.,  
 Fourier Analysis, Fuzzy Sets and Systems

23) Rainer Nagel  
 Arbeitsbereich Funktionalanalysis  
 Mathematisches Institut  
 Auf der Morgenstelle 10  
 D-72076 Tuebingen  
 Germany  
 tel. 49-7071-2973242  
 fax 49-7071-294322  
[rana@fa.uni-tuebingen.de](mailto:rana@fa.uni-tuebingen.de)  
 evolution equations, semigroups, spectral th.,  
 positivity

24) Panos M. Pardalos  
 Center for Appl. Optimization  
 University of Florida  
 303 Weil Hall  
 P.O. Box 116595  
 Gainesville, FL 32611-6595  
 tel. 352-392-9011  
[pardalos@ufl.edu](mailto:pardalos@ufl.edu)  
 Optimization, Operations Research



## BELGIUM

f.bastin@ulg.ac.be  
Functional Analysis, Wavelets

5) Yeol Je Cho  
Department of Mathematics Education  
College of Education  
Gyeongsang National University  
Chinju 660-701

## KOREA

tel. 055-751-5673 Office,  
055-755-3644 home,  
fax 055-751-6117  
yjcho@nongae.gsnu.ac.kr  
Nonlinear operator Th., Inequalities,  
Geometry of Banach Spaces

6) Sever S. Dragomir  
School of Communications and Informatics  
Victoria University of Technology  
PO Box 14428  
Melbourne City M.C  
Victoria 8001, Australia  
tel 61 3 9688 4437, fax 61 3 9688 4050  
sever.dragomir@vu.edu.au,  
sever@sci.vu.edu.au  
Math. Analysis, Inequalities, Approximation  
Th.,  
Numerical Analysis, Geometry of Banach  
Spaces,  
Information Th. and Coding

7) Angelo Favini  
Università di Bologna  
Dipartimento di Matematica  
Piazza di Porta San Donato 5  
40126 Bologna, ITALY  
tel. ++39 051 2094451  
fax. ++39 051 2094490  
favini@dm.unibo.it  
Partial Differential Equations, Control  
Theory,  
Differential Equations in Banach Spaces

8) Claudio A. Fernandez  
Facultad de Matematicas  
Pontificia Universidad Católica de Chile  
Vicuna Mackenna 4860  
Santiago, Chile  
tel. ++56 2 354 5922  
fax. ++56 2 552 5916  
cfernand@mat.puc.cl  
Partial Differential Equations,  
Mathematical Physics,  
Scattering and Spectral Theory

25) Svetlozar T. Rachev  
Dept. of Statistics and Applied Probability  
Program

University of California, Santa Barbara  
CA 93106-3110, USA  
tel. 805-893-4869  
rachev@pstat.ucsb.edu

## AND

Chair of Econometrics and Statistics  
School of Economics and Business Engineering  
University of Karlsruhe  
Kollegium am Schloss, Bau II, 20.12, R210  
Postfach 6980, D-76128, Karlsruhe, Germany  
tel. 011-49-721-608-7535  
rachev@lsoe.uni-karlsruhe.de  
Mathematical and Empirical Finance,  
Applied Probability, Statistics and Econometrics

26) John Michael Rassias  
University of Athens  
Pedagogical Department  
Section of Mathematics and Informatics  
20, Hippocratous Str., Athens, 106 80, Greece

Address for Correspondence

4, Agamemnonos Str.  
Aghia Paraskevi, Athens, Attikis 15342 Greece  
jrassias@primedu.uoa.gr  
jrassias@tellas.gr  
Approximation Theory, Functional Equations,  
Inequalities, PDE

27) Paolo Emilio Ricci  
Università degli Studi di Roma "La Sapienza"  
Dipartimento di Matematica-Istituto  
"G. Castelnuovo"  
P.le A. Moro, 2-00185 Roma, ITALY  
tel. ++39 0649913201, fax ++39 0644701007  
riccip@uniroma1.it, Paoloemilio.Ricci@uniroma1.it  
Orthogonal Polynomials and Special functions,  
Numerical Analysis, Transforms, Operational  
Calculus,  
Differential and Difference equations

28) Cecil C. Rousseau  
Department of Mathematical Sciences  
The University of Memphis  
Memphis, TN 38152, USA  
tel. 901-678-2490, fax 901-678-2480  
ccrousse@memphis.edu  
Combinatorics, Graph Th.,  
Asymptotic Approximations,  
Applications to Physics

29) Tomasz Rychlik

- 9) A.M.Fink  
Department of Mathematics  
Iowa State University  
Ames, IA 50011-0001, USA  
tel.515-294-8150  
fink@math.iastate.edu  
Inequalities, Ordinary Differential Equations
- 10) Sorin Gal  
Department of Mathematics  
University of Oradea  
Str.Armatei Romane 5  
3700 Oradea, Romania  
galso@uoradea.ro  
Approximation Th., Fuzzyness, Complex Analysis
- 11) Jerome A.Goldstein  
Department of Mathematical Sciences  
The University of Memphis,  
Memphis, TN 38152, USA  
tel.901-678-2484  
jgoldste@memphis.edu  
Partial Differential Equations, Semigroups of Operators
- 12) Heiner H.Gonska  
Department of Mathematics  
University of Duisburg  
Duisburg, D-47048  
Germany  
tel.0049-203-379-3542 office  
gonska@informatik.uni-duisburg.de  
Approximation Th., Computer Aided Geometric Design
- 13) Dmitry Khavinson  
Department of Mathematical Sciences  
University of Arkansas  
Fayetteville, AR 72701, USA  
tel.(479)575-6331, fax(479)575-8630  
dmitry@uark.edu  
Potential Th., Complex Analysis, Holomorphic PDE,  
Approximation Th., Function Th.
- 14) Virginia S.Kiryakova  
Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
virginia@diogenes.bg  
Special Functions, Integral Transforms, Fractional Calculus
- 15) Hans-Bernd Knoop  
Institute of Mathematics  
Polish Academy of Sciences  
Chopina 12, 87100 Torun, Poland  
T.Rychlik@impan.gov.pl  
Mathematical Statistics, Probabilistic Inequalities
- 30) Bl. Sendov  
Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
bsendov@bas.bg  
Approximation Th., Geometry of Polynomials, Image Compression
- 31) Igor Shevchuk  
Faculty of Mathematics and Mechanics  
National Taras Shevchenko  
University of Kyiv  
252017 Kyiv  
UKRAINE  
shevchuk@univ.kiev.ua  
Approximation Theory
- 32) H.M.Srivastava  
Department of Mathematics and Statistics  
University of Victoria  
Victoria, British Columbia V8W 3P4  
Canada  
tel.250-721-7455 office, 250-477-6960 home,  
fax 250-721-8962  
harimsri@math.uvic.ca  
Real and Complex Analysis, Fractional Calculus and Appl.,  
Integral Equations and Transforms, Higher Transcendental Functions and Appl., q-Series and q-Polynomials, Analytic Number Th.
- 33) Stevo Stevic  
Mathematical Institute of the Serbian Acad. of Science  
Knez Mihailova 35/I  
11000 Beograd, Serbia  
sstevic@ptt.yu; sstevo@matf.bg.ac.yu  
Complex Variables, Difference Equations, Approximation Th., Inequalities
- 34) Ferenc Szidarovszky  
Dept. Systems and Industrial Engineering  
The University of Arizona  
Engineering Building, 111  
PO.Box 210020  
Tucson, AZ 85721-0020, USA  
szidar@sie.arizona.edu  
Numerical Methods, Game Th., Dynamic Systems,

Institute of Mathematics  
 Gerhard Mercator University  
 D-47048 Duisburg  
 Germany  
 tel.0049-203-379-2676  
 knoop@math.uni-duisburg.de  
 Approximation Theory, Interpolation

16) Jerry Koliha  
 Dept. of Mathematics & Statistics  
 University of Melbourne  
 VIC 3010, Melbourne  
 Australia  
 koliha@unimelb.edu.au  
 Inequalities, Operator Theory,  
 Matrix Analysis, Generalized Inverses

17) Mustafa Kulenovic  
 Department of Mathematics  
 University of Rhode Island  
 Kingston, RI 02881, USA  
 kulenm@math.uri.edu  
 Differential and Difference Equations

18) Gerassimos Ladas  
 Department of Mathematics  
 University of Rhode Island  
 Kingston, RI 02881, USA  
 gladas@math.uri.edu  
 Differential and Difference Equations

19) V. Lakshmikantham  
 Department of Mathematical Sciences  
 Florida Institute of Technology  
 Melbourne, FL 32901  
 e-mail: lakshmik@fit.edu  
 Ordinary and Partial Differential  
 Equations,  
 Hybrid Systems, Nonlinear Analysis

20) Rupert Lasser  
 Institut für Biomathematik & Biomertie, GSF  
 -National Research Center for environment  
 and health  
 Ingolstaedter landstr.1  
 D-85764 Neuherberg, Germany  
 lasser@gsf.de  
 Orthogonal Polynomials, Fourier Analysis,  
 Mathematical Biology

Multicriteria Decision making,  
 Conflict Resolution, Applications  
 in Economics and Natural Resources  
 Management

35) Gancho Tachev  
 Dept. of Mathematics  
 Univ. of Architecture, Civil Eng. and Geodesy  
 1 Hr. Smirnenski blvd  
 BG-1421 Sofia, Bulgaria  
 gtt\_fte@uacg.bg  
 Approximation Theory

36) Manfred Tasche  
 Department of Mathematics  
 University of Rostock  
 D-18051 Rostock  
 Germany  
 manfred.tasche@mathematik.uni-rostock.de  
 Approximation Th., Wavelet, Fourier Analysis,  
 Numerical Methods, Signal Processing,  
 Image Processing, Harmonic Analysis

37) Chris P. Tsokos  
 Department of Mathematics  
 University of South Florida  
 4202 E. Fowler Ave., PHY 114  
 Tampa, FL 33620-5700, USA  
 profcpt@math.usf.edu, profcpt@chumal.cas.usf.edu  
 Stochastic Systems, Biomathematics,  
 Environmental Systems, Reliability Th.

38) Lutz Volkmann  
 Lehrstuhl II für Mathematik  
 RWTH-Aachen  
 Templergraben 55  
 D-52062 Aachen  
 Germany  
 volkm@math2.rwth-aachen.de  
 Complex Analysis, Combinatorics, Graph Theory

**EDITOR'S NOTE**

This special issue on “Applied Mathematics and Approximation Theory” contains expanded versions of articles that were presented in the international conference “Applied Mathematics and Approximation Theory 2008” ( AMAT 08), during October 11-13, 2008 at the University of Memphis, Memphis, Tennessee, USA.

All articles were refereed.

The organizer and Editor

George Anastassiou

# Approximation by Nonlinear Bernstein and Favard-Szász-Mirakjan Operators of Max-Product Kind

Barnabás Bede<sup>1</sup> and Sorin G. Gal<sup>2</sup>

<sup>1</sup>Department of Mathematics,  
The University of Texas-Pan American,  
1201 West University, Edinburg, Tx, 78539, USA  
E-mail: bedeb@utpa.edu, bede.barna@bgk.bmf.hu

<sup>2</sup>Department of Mathematics and Computer Science,  
The University of Oradea,  
Universitatii 1, 410087, Oradea, Romania  
E-mail: galso@uoradea.ro

## Abstract

We address in the present paper the following problem : is the linear structure the only one which allows us to construct approximation operators ? As an answer to this problem we propose new, so-called pseudo-linear approximation operators, which are defined in a max-product algebra. We consider a nonlinear (max-product) Bernstein operator and a nonlinear (max-product) Favard-Szász-Mirakjan operator. We prove that the approximation errors given by these operators are similar to those of the corresponding linear operators provided by the classical Approximation Theory. Also, by some graphs we illustrate that the nonlinear operators can better reproduce the shape of some continuous functions (which are not differentiable or have abrupt changes of the values in some points) than their linear counterparts.

## 1 Introduction

The linear approximation operators provided by the classical Approximation Theory use exclusively as underlying algebraic structure the linear space structure of the reals. In the present paper we address the following problem : is the linear structure the only one which allows us to construct approximation operators ?

This problem was proposed in [1], [2] for Shepard operators. The findings in these two papers were that besides the linear structure we may opt for different structures as max-product, max-min (fuzzy) algebras or semirings with generated pseudo-operations. In this sense, Shepard-type operators of Max-Product and Max-Min kinds were studied together with operators of Shepard-type based on generated pseudo-operations. Note that these operators are nonlinear (they are so-called pseudo-linear). So, the topic of these papers and the present one partially fits on the area of Nonlinear Approximation, as it is described by e.g. [3], [5]. It is only a partial fit on this area, because while the operators are nonlinear, the algebraic structure underlying them is still a structure of linear space. In [3] and [5], the nonlinearity comes from the construction of approximations based on a dictionary of functions. Because of this reason, the method is not fully constructive. In contrast, the approach presented here is fully constructive and quite simple.

Recently, the monograph [4] (see pp. 324-326, Open Problem 5.5.4) brings into attention the same question of constructing approximation operators with max-product or max-min operations, but in a more systematic way. Thus, following [4] we consider the nonlinear Bernstein operator of max-product kind

$$B_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^n p_{n,k}(x) f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^n p_{n,k}(x)},$$

with  $p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}$  and the nonlinear Favard-Szász-Mirakjan operator of max-product kind

$$F_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!} f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!}}.$$

Our findings are that these operators have very similar properties to the corresponding linear operators provided by the classical Approximation Theory. In this sense we show that the error estimates in approximation by these operators is of order  $\mathcal{O}\left[\omega_1\left(f; \frac{1}{\sqrt{n}}\right)\right]$ . Some experimental results are also discussed and our finding is that the proposed max-product operators can better follow abrupt changes in the target function than their classical linear counterparts.

## 2 Nonlinear approximation operators

Over the set of positive reals,  $\mathbb{R}_+$ , we consider the operations  $\vee$  (maximum) and  $\cdot$  product. Then  $(\mathbb{R}_+, \vee, \cdot)$  has a semiring structure and we call it as Max-Product algebra.

Let  $I \subset \mathbb{R}$  be a bounded or unbounded interval, and

$$CB_+(I) = \{f : I \rightarrow \mathbb{R}_+; f \text{ continuous and bounded on } I\}.$$

The general discrete form of a max-product approximation operator  $L_n : CB_+(I) \rightarrow CB_+(I)$ , is

$$L_n(f)(x) = \bigvee_{i=0}^n K_n(x, x_i) \cdot f(x_i),$$

or

$$L_n(f)(x) = \bigvee_{i=0}^{\infty} K_n(x, x_i) \cdot f(x_i),$$

where  $n \in \mathbb{N}$ ,  $f \in CB_+(I)$ ,  $K_n(\cdot, x_i) \in CB_+(I)$  and  $x_i \in I$ , for all  $i$ . These operators are nonlinear, positive operators and moreover they satisfy a pseudo-linearity condition of the form

$$L_n(\alpha \cdot f \vee \beta \cdot g)(x) = \alpha \cdot L_n(f)(x) \vee \beta \cdot L_n(g)(x), \forall \alpha, \beta \in \mathbb{R}_+, f, g : I \rightarrow \mathbb{R}_+.$$

In this section we present some general results on positive nonlinear operators. These are used later in the study of nonlinear Bernstein and Favard-Szász-Mirakjan operators of max-product kind.

**Lemma 1** *Let  $I \subset \mathbb{R}$  be a bounded or unbounded interval,*

$$CB_+(I) = \{f : I \rightarrow \mathbb{R}_+; f \text{ continuous and bounded on } I\},$$

*and  $L_n : CB_+(I) \rightarrow CB_+(I)$ ,  $n \in \mathbb{N}$  be a sequence of operators satisfying the following properties :*

- (i) *if  $f, g \in CB_+(I)$  satisfy  $f \leq g$  then  $L_n(f) \leq L_n(g)$  for all  $n \in \mathbb{N}$  ;*
  - (ii)  *$L_n(f + g) \leq L_n(f) + L_n(g)$  for all  $f, g \in CB_+(I)$ .*
- Then for all  $f, g \in CB_+(I)$ ,  $n \in \mathbb{N}$  and  $x \in I$  we have*

$$|L_n(f)(x) - L_n(g)(x)| \leq L_n(|f - g|)(x).$$

**Proof.** Let  $f, g \in CB_+(I)$ . We have  $f = f - g + g \leq |f - g| + g$ , which by the conditions (i) – (ii) successively implies  $L_n(f)(x) \leq L_n(|f - g|)(x) + L_n(g)(x)$ , that is  $L_n(f)(x) - L_n(g)(x) \leq L_n(|f - g|)(x)$ .

Writing now  $g = g - f + f \leq |f - g| + f$  and applying the above reasonings, it follows  $L_n(g)(x) - L_n(f)(x) \leq L_n(|f - g|)(x)$ , which combined with the above inequality gives  $|L_n(f)(x) - L_n(g)(x)| \leq L_n(|f - g|)(x)$ . ■

**Remark 2** 1) *It is easy to see that our max-product operators satisfy the conditions in Lemma 1, (i), (ii). In fact, instead of (i) they satisfy the stronger condition*

$$L_n(f \vee g)(x) = L_n(f)(x) \vee L_n(g)(x), \quad f, g \in CB_+(I).$$

*Indeed, taking in the above equality  $f \leq g$ ,  $f, g \in CB_+(I)$ , it easily follows  $L_n(f)(x) \leq L_n(g)(x)$ .*

2) *In addition, it is immediate that the max-product operators are positive homogenous, that is  $L_n(\lambda f) = \lambda L_n(f)$  for all  $\lambda \geq 0$ .*

**Corollary 3** *Let  $L_n : CB_+(I) \rightarrow CB_+(I)$ ,  $n \in N$  be a sequence of operators satisfying the conditions (i)-(ii) in Lemma 1 and in addition being positive homogenous. Then for all  $f \in CB_+(I)$ ,  $n \in N$  and  $x \in I$  we have*

$$|f(x) - L_n(x)| \leq \left[ \frac{1}{\delta} L_n(\varphi_x)(x) + L_n(e_0)(x) \right] \omega_1(f; \delta)_I + f(x) \cdot |L_n(e_0)(x) - 1|,$$

where  $\delta > 0$ ,  $e_0(t) = 1$  for all  $t \in I$ ,  $\varphi_x(t) = |t - x|$  for all  $t \in I$ ,  $x \in I$  and  $\omega_1(f; \delta)_I = \max\{|f(x) - f(y)|; x, y \in I, |x - y| \leq \delta\}$ .

**Proof.** The proof is identical with that for positive linear operators. Indeed, from the identity

$$L_n(f)(x) - f(x) = [L_n(f)(x) - f(x) \cdot L_n(e_0)(x)] + f(x)[L_n(e_0)(x) - 1],$$

it follows (by the positive homogeneity and by Lemma 1)

$$\begin{aligned} |f(x) - L_n(f)(x)| &\leq |L_n(f)(x) - L_n(f(t))(x)| + |f(x)| \cdot |L_n(e_0)(x) - 1| \leq \\ &L_n(|f(t) - f(x)|)(x) + |f(x)| \cdot |L_n(e_0)(x) - 1|. \end{aligned}$$

Now, since for all  $t, x \in I$  we have

$$|f(t) - f(x)| \leq \omega_1(f; |t - x|)_I \leq \left[ \frac{1}{\delta} |t - x| + 1 \right] \omega_1(f; \delta)_I,$$

replacing above we immediately obtain the estimate in the statement. ■

An immediate consequence of Corollary 3 is the following.

**Corollary 4** *Suppose that in addition to the conditions in Corollary 3, the sequence  $(L_n)_n$  satisfies  $L_n(e_0) = e_0$ , for all  $n \in N$ . Then for all  $f \in CB_+(I)$ ,  $n \in N$  and  $x \in I$  we have*

$$|f(x) - L_n(x)| \leq \left[ 1 + \frac{1}{\delta} L_n(\varphi_x)(x) \right] \omega_1(f; \delta)_I.$$

**Remark 5** *The max-product operators that we will construct in the following sections will satisfy the additional condition in Corollary 3, so that Corollary 3 one applies to these kind of operators.*

### 3 Nonlinear Bernstein Operator

Let  $f : [0, 1] \rightarrow \mathbb{R}_+$  be continuous. We consider the following nonlinear Bernstein operator of max-product type

$$B_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^n p_{n,k}(x) f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^n p_{n,k}(x)}, \quad (1)$$



with

$$p_{n,k}(x) = \binom{n}{k} x^k (1-x)^{n-k}.$$

We provide an error estimate for this approximation operator in terms of the modulus of continuity.

**Theorem 6** *The following pointwise estimate holds true for the nonlinear Bernstein operator*

$$|B_n^{(M)}(f)(x) - f(x)| \leq C \omega_1 \left( f, \frac{\sqrt{x(1-x)}}{\sqrt{n}} + \frac{1}{n} \right)_{[0,1]}, \quad x \in [0, 1], n \in \mathbb{N},$$

where  $C > 0$  is an absolute constant independent of  $f$ ,  $n$  and  $x$ .

$$\omega_1(f, \delta)_{[0,1]} = \sup\{|f(x) - f(y)|; x, y \in [0, 1], |x - y| \leq \delta\}.$$

**Proof.** It is easy to check that the max-product Bernstein operators fulfil the conditions in Corollary 4 and we have

$$|B_n^{(M)}(f)(x) - f(x)| \leq \left( 1 + \frac{1}{\delta_n} B_n^{(M)}(\varphi_x)(x) \right) \omega_1(f, \delta_n)_{[0,1]},$$

where  $\varphi_x(t) = |t - x|$ . So, it is enough to estimate

$$B_n^{(M)}(\varphi_x)(x) = \frac{\bigvee_{k=0}^n p_{n,k}(x) \left| \frac{k}{n} - x \right|}{\bigvee_{k=0}^n p_{n,k}(x)}.$$

First we calculate  $\bigvee_{k=0}^n p_{n,k}(x)$  for fixed  $n, x$ . Let  $E_{n,k}(x) = p_{n,k+1}(x) - p_{n,k}(x)$ ,  $k \in \{0, \dots, n-1\}$ . We have by successive calculations

$$\begin{aligned} E_{n,k}(x) &= \binom{n}{k+1} x^{k+1} (1-x)^{n-k-1} - \binom{n}{k} x^k (1-x)^{n-k} \\ &= \binom{n}{k} x^k (1-x)^{n-k-1} \frac{nx - k - 1 + x}{k+1}. \end{aligned}$$

We have  $nx - k - 1 + x \geq 0$  if and only if  $E_{n,k}(x) \geq 0$ , i.e.,  $p_{n,k+1}(x) \geq p_{n,k}(x)$ . Further  $nx - k - 1 + x \geq 0$  if  $k \leq nx - (1-x)$ , i.e.,  $k \leq [(n+1)x] - 1$ . It follows

that  $\bigvee_{k=0}^n p_{n,k}(x) = p_{n,r}(x) = \binom{n}{r} x^r (1-x)^{n-r}$ , with  $r = [(n+1)x]$ .

Now we will estimate the ratio

$$\frac{p_{n,k}(x) \left| \frac{k}{n} - x \right|}{p_{n,r}(x)}, \quad r = [(n+1)x].$$

We observe that  $nx < nx + x < nx + 1$ , so we have two cases:  $r = [nx]$  or  $r = [nx] + 1$ . We will present only the case  $r = [nx]$ , similar result being true for  $r = [nx] + 1$ .

We observe that  $r = [nx]$  implies  $nx - 1 \leq r < nx$ , i.e.,

$$\frac{r}{n} < x \leq \frac{r+1}{n}. \quad (2)$$

Case 1. If  $k \leq r$  then we have

$$\left| \frac{k}{n} - x \right| = x - \frac{k}{n} \leq \frac{r-k+1}{n}. \quad (3)$$

We estimate in what follows the expression

$$A_{n,k,r} = \frac{p_{n,k}(x) \left| \frac{k}{n} - x \right|}{p_{n,r}(x)}, k \leq r.$$

By (3) we have

$$\begin{aligned} A_{n,k,r} &= \frac{\binom{n}{k} x^k (1-x)^{n-k} \left| \frac{k}{n} - x \right|}{\binom{n}{r} x^r (1-x)^{n-r}} \\ &\leq \frac{r!(n-r)!}{k!(n-k)!} \left( \frac{1-x}{x} \right)^{r-k} \frac{r-k+1}{n}. \end{aligned}$$

Since the function  $\frac{1-x}{x}$  is decreasing for  $x \in [0, 1]$  we have by (2)  $\frac{1-x}{x} \leq \frac{n-r}{r}$ . Further, we have

$$A_{n,k,r} \leq \frac{r!(n-r)!}{k!(n-k)!} \left( \frac{n-r}{r} \right)^{r-k} \frac{r-k+1}{n} = B_{r,k,n}$$

We will find the maximal term  $B_{r,k,n}$  for fixed  $n, r$  and variable  $k \leq r$ . We have

$$\begin{aligned} B_{r,k+1,n} - B_{r,k,n} &= \frac{1}{n} \frac{r!(n-r)!}{k!(n-k-1)!} \left( \frac{n-r}{r} \right)^{r-k-1} \left( \frac{r-k}{k+1} - \frac{(n-r)(r-k+1)}{r(n-k)} \right) \\ &= \frac{r!(n-r)!}{k!(n-k-1)!} \left( \frac{n-r}{r} \right)^{r-k-1} \frac{(r-k)^2 - r - 1 + \frac{r^2+r}{n}}{r(n-k)(k+1)}. \end{aligned}$$

We observe that  $(r-k)^2 - r - 1 + \frac{r^2+r}{n} > 0$  for  $(r-k)^2 > r+1 - \frac{r^2+r}{n}$ , i.e.,

$r-k > \sqrt{r+1 - \frac{r^2+r}{n}}$  and this holds for  $k < r - \sqrt{\frac{(r+1)(n-r)}{n}}$ . The maximal term, therefore is given by

$$k_0 = \left\lceil r - \sqrt{\frac{(r+1)(n-r)}{n}} \right\rceil + 1 = r - \left\lfloor \sqrt{\frac{(r+1)(n-r)}{n}} \right\rfloor,$$

so we have  $A_{nkr} \leq B_{n,k,r} \leq B_{n,k_0,r}$ . Further, by Stirling formula we obtain

$$\begin{aligned} B_{n,k_0,r} &\sim \frac{\frac{\sqrt{2\pi r}}{e^r} r^r \frac{\sqrt{2\pi(n-r)}}{e^{n-r}} (n-r)^{n-r}}{\frac{\sqrt{2\pi k_0}}{e^{k_0}} k_0^{k_0} \frac{\sqrt{2\pi(n-k_0)}}{e^{n-k_0}} (n-k_0)^{n-k_0}} \left( \frac{n-r}{r} \right)^{r-k_0} \frac{r-k_0+1}{n} \\ &= \left( \frac{r}{k_0} \right)^{k_0+\frac{1}{2}} \left( \frac{n-r}{n-k_0} \right)^{n-k_0+\frac{1}{2}} (r-k_0+1) \frac{1}{n}. \end{aligned}$$

Now, since  $k_0 \sim r - \sqrt{\frac{(r+1)(n-r)}{n}}$  we get

$$B_{n,k_0,r} \sim \sqrt{\frac{r}{r - \sqrt{\frac{(r+1)(n-r)}{n}}} \frac{n-r}{n-r + \sqrt{\frac{(r+1)(n-r)}{n}}}} \left( \sqrt{\frac{(r+1)(n-r)}{n}} + 1 \right) \frac{1}{n}.$$

We observe that

$$\frac{r}{r - \sqrt{\frac{(r+1)(n-r)}{n}}} \frac{n-r}{n-r + \sqrt{\frac{(r+1)(n-r)}{n}}} = \frac{1}{1 + \sqrt{\frac{(r+1)}{nr^2(n-r)}} (2r+1) - \frac{(r+1)}{nr}}$$

and

$$\sqrt{\frac{(r+1)}{nr^2(n-r)}} (2r+1) \geq 2\sqrt{\frac{r}{n(n-r)}}$$

so we obtain

$$\frac{1}{1 + \sqrt{\frac{(r+1)}{nr^2(n-r)}} (2r+1) - \frac{(r+1)}{nr}} \leq \frac{1}{1 + 2\sqrt{\frac{r}{n(n-r)}} - \frac{1}{n}} \sim \mathcal{O}(1).$$

Finally, we obtain

$$A_{n,k,r} \leq C \left( \sqrt{\frac{(r+1)(n-r)}{n}} + 1 \right) \frac{1}{n}.$$

Since  $r = [nx]$  we get

$$A_{n,k,r} \leq C \left( \sqrt{\frac{(nx+1)(n-nx)}{n}} + 1 \right) \frac{1}{n} \sim C \left( \frac{\sqrt{x(1-x)}}{\sqrt{n}} + \frac{1}{n} \right).$$

Case 2. If  $k > r$  then

$$\left| \frac{k}{n} - x \right| = \frac{k}{n} - x \leq \frac{k-r}{n}.$$

We estimate the expression

$$A_{n,k,r} = \frac{p_{n,k}(x) \left| \frac{k}{n} - x \right|}{p_{n,r}(x)}, k > r.$$

We have

$$A_{n,k,r} = \frac{\binom{n}{k} x^k (1-x)^{n-k} \left| \frac{k}{n} - x \right|}{\binom{n}{r} x^r (1-x)^{n-r}} \leq \frac{r!(n-r)!}{k!(n-k)!} \left( \frac{1-x}{x} \right)^{r-k} \frac{k-r}{n}.$$

Since the function  $\frac{1-x}{x}$  is decreasing we have by (2)  $\frac{1-x}{x} \geq \frac{n-r-1}{r+1}$ . Further we obtain

$$A_{n,k,r} \leq \frac{r!(n-r)!}{k!(n-k)!} \left( \frac{r+1}{n-r-1} \right)^{k-r} \frac{k-r}{n} = B_{r,k,n}.$$

We find, similar to Case 1, the maximal term  $B_{r,k,n}$  for fixed  $n, r$  to be obtained when  $k_0 = \left\lceil r + \sqrt{\frac{(r+1)(n-r)}{n}} \right\rceil + 1 = r + \left\lceil \sqrt{\frac{(r+1)(n-r)}{n}} \right\rceil$ . Now we have  $k_0 \sim r + \sqrt{\frac{(r+1)(n-r)}{n}}$  and so,

$$B_{n,k_0,r} \sim \sqrt{\frac{r}{r - \sqrt{\frac{(r+1)(n-r)}{n}}} \frac{n-r}{n-r + \sqrt{\frac{(r+1)(n-r)}{n}}}} \left( \sqrt{\frac{(r+1)(n-r)}{n}} + 1 \right) \frac{1}{n}.$$

Similar to case 1, finally we obtain

$$A_{n,k,r} \leq C \left( \sqrt{\frac{(r+1)(n-r)}{n}} + 1 \right) \frac{1}{n}.$$

and

$$A_{n,k,r} \leq C \left( \sqrt{\frac{(nx+1)(n-nx)}{n}} + 1 \right) \frac{1}{n} \sim C \left( \frac{\sqrt{x(1-x)}}{\sqrt{n}} + \frac{1}{n} \right).$$

Taking into account the estimates in Cases 1 and 2 we get

$$B_n^{(M)}(\varphi_x)(x) \leq C \left( \frac{\sqrt{x(1-x)}}{\sqrt{n}} + \frac{1}{n} \right) \leq \frac{C_1}{\sqrt{n}}.$$

Taking  $\delta_n = \frac{\sqrt{x(1-x)}}{\sqrt{n}} + \frac{1}{n}$  we obtain

$$|B_n^{(M)}(f)(x) - f(x)| \leq C\omega_1 \left( f, \frac{\sqrt{x(1-x)}}{\sqrt{n}} + \frac{1}{n} \right)_{[0,1]},$$

which completes the proof. ■

**Remark 7** 1) The error estimate shown by the above theorem is similar to the error estimate for the linear Bernstein operators. However, since obviously  $B_n^{(M)}(f)(0) = f(0)$  and  $B_n^{(M)}(f)(1) = f(1)$ , it is clear that the term  $1/n$  in the estimate of Theorem 6 could be dropped.

2) The proof of the above theorem evidently implies the uniform estimate

$$\max_{x \in [0,1]} |B_n^{(M)}(f)(x) - f(x)| \leq C_1 \omega_1 \left( f, \frac{1}{\sqrt{n}} \right)_{[0,1]}, n \in \mathbb{N}.$$

However, we conjecture that in fact the order of approximation  $\mathcal{O}[\omega_1(f; 1/\sqrt{n})]$  in Theorem 6 might be improved to the order  $\mathcal{O}[\omega_1(f; \ln(n)/n)]$ .

## 4 Nonlinear Favard-Szász-Mirakjan Operator

Let  $f : [0, \infty) \rightarrow \mathbb{R}_+$  be continuous. We consider the following nonlinear Favard-Szász-Mirakjan operator

$$F_n^{(M)}(f)(x) = \frac{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!} f\left(\frac{k}{n}\right)}{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!}},$$

We provide an error estimate for this approximation operator in terms of the modulus of continuity similarly to the case of Bernstein operators.

**Theorem 8** *The following pointwise estimate holds true for the nonlinear Favard-Szász-Mirakjan operator*

$$|F_n^{(M)}(f)(x) - f(x)| \leq C \omega_1 \left( f, \frac{\sqrt{x}}{\sqrt{n}} \right)_{[0,\infty)}, x \geq 0, n \in \mathbb{N},$$

where  $C > 0$  is an absolute constant (that is independent of  $f$ ,  $n$  and  $x$ ) and

$$\omega_1(f, \delta)_{[0,\infty)} = \sup\{|f(x) - f(y)|; x, y \geq 0, |x - y| \leq \delta\}.$$

**Proof.** By Corollary 4, the error bound is controlled by the ratio

$$B_n^{(M)}(\varphi_x)(x) = \frac{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!} \left| \frac{k}{n} - x \right|}{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!}}.$$

First we calculate  $\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!}$  for fixed  $n \in \mathbb{N}, x \in [0, \infty)$ . We observe that

$$E_{n,k}(x) = \frac{(nx)^{k+1}}{(k+1)!} - \frac{(nx)^k}{k!} = \frac{(nx)^k}{k!} \frac{nx - k - 1}{k+1}.$$

We have  $E_{n,k}(x) \geq 0$  if and only if  $k \leq nx - 1$ . It follows that  $\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!} = \frac{(nx)^r}{r!}$ , with  $r = [nx]$ .

Now we estimate the ratio

$$R_{n,r}(x) = \frac{\bigvee_{k=0}^{\infty} \frac{(nx)^k}{k!} \left| \frac{k}{n} - x \right|}{\frac{(nx)^r}{r!}}, r = [nx].$$

Case 1. If  $k \leq r$  then we have

$$\left| \frac{k}{n} - x \right| = x - \frac{k}{n} \leq \frac{r - k + 1}{n}.$$

We estimate in what follows the expression

$$A_{n,k,r} = \frac{\frac{(nx)^k}{k!} \left| \frac{k}{n} - x \right|}{\frac{(nx)^r}{r!}}, k \leq r.$$

As in the proof of the estimate for the Bernstein operator, we have by (3) and (2)

$$A_{n,k,r} \leq \frac{(nx)^k}{k!} \frac{r!}{(nx)^r} \frac{r - k + 1}{n} \leq r^{k-r} \frac{r!}{k!} \frac{r - k + 1}{n}$$

Let  $E_{r,k} = r^{k-r} \frac{r!}{k!} (r - k + 1)$ . We have

$$\begin{aligned} E_{r,k+1} - E_{r,k} &= r^{k-r+1} \frac{r!}{(k+1)!} (r - k) - r^{k-r} \frac{r!}{k!} (r - k + 1) \\ &= r^{k-r} \frac{r!}{k!} \frac{(r - k)^2 - r - 1}{k + 1} \end{aligned}$$

It is easy to check that the maximal term  $E_{r,k}$  is attained for  $k_0 = [r + 1 - \sqrt{r + 1}] = r - [\sqrt{r + 1}]$ .

By Stirling formula we obtain

$$\begin{aligned} E_{r,k} &\leq E_{r,k_0} = r^{k_0-r} \frac{\sqrt{2\pi r} \frac{r^r}{e^r}}{\sqrt{2\pi k_0} \frac{k_0^{k_0}}{e^{k_0}}} (r - k_0 + 1) \\ &= \frac{r^{r-[\sqrt{r+1}]+\frac{1}{2}}}{(r - [\sqrt{r+1}])^{r-[\sqrt{r+1}]+\frac{1}{2}}} e^{-[\sqrt{r+1}]} ([\sqrt{r+1}] + 1). \end{aligned}$$

Further we have,

$$\begin{aligned} E_{r,k} &= \frac{r^{r-\sqrt{r+1}+\frac{1}{2}}}{(r - \sqrt{r+1})^{r-\sqrt{r+1}+\frac{1}{2}}} e^{-\sqrt{r+1}} (\sqrt{r+1} + 1) \\ &\sim \sqrt{\frac{r}{r - \sqrt{r+1}}} (\sqrt{r+1} + 1) = \sqrt{\frac{r}{r - \sqrt{r+1}}} (\sqrt{r+1} + 1), \end{aligned}$$

and taking into account that

$$\sqrt{\frac{r}{r - \sqrt{r+1}}}(\sqrt{r+1} + 1) \leq (\sqrt{r+1} + 1) \leq (\sqrt{nx+1} + 1)$$

finally we obtain

$$A_{n,k,r} \leq \frac{\sqrt{nx+1} + 1}{n} = \mathcal{O}\left(\frac{\sqrt{x}}{\sqrt{n}}\right).$$

Case 2. If  $k > r$  then we have

$$A_{n,k,r} = \frac{\frac{(nx)^k}{k!} \left| \frac{k}{n} - x \right|}{\frac{(nx)^r}{r!}} \leq (r+1)^{k-r} \frac{r!}{k!} \frac{k-r}{n}.$$

Further, if  $E_{r,k} = (r+1)^{k-r} \frac{r!}{k!} (k-r)$ , similar to Case 1, the maximal term is attained for  $k_0 = r + \lceil \sqrt{r+1} \rceil$ .

By Stirling formula we have by successive calculations

$$E_{r,k} \leq E_{r,k_0} \sim \sqrt{\frac{r}{r + \sqrt{r+1}}} (\sqrt{r+1} + 1).$$

Further we have  $\sqrt{\frac{r}{r + \sqrt{r+1}}}(\sqrt{r+1} + 1) \leq (\sqrt{nx+1} + 1)$  and finally we obtain

$$A_{n,k,r} \leq \frac{\sqrt{nx+1} + 1}{n} = \mathcal{O}\left(\frac{\sqrt{x}}{\sqrt{n}}\right), k > r.$$

Taking into account the estimates in cases 1 and 2 we get

$$B_n^{(M)}(\varphi_x)(x) = \mathcal{O}\left(\frac{\sqrt{x}}{\sqrt{n}}\right).$$

Taking  $\delta_n = \frac{\sqrt{x}}{\sqrt{n}} + \frac{1}{n}$  we obtain

$$|F_n^{(M)}(f)(x) - f(x)| \leq C\omega_1\left(f, \frac{\sqrt{x}}{\sqrt{n}}\right)_{[0,\infty)}.$$

■

**Remark 9** 1) The error estimate shown in this theorem is similar to that for the linear Favard-Szász-Mirkjan operator.

2) The pointwise error estimate in the above theorem obviously implies the uniform estimate

$$\max_{x \in [0,a]} |F_n^{(M)}(f)(x) - f(x)| \leq C\omega_1\left(f, \frac{1}{\sqrt{n}}\right)_{[0,\infty)},$$

for any  $a > 0$ . However, we conjecture that in fact the order of approximation  $\mathcal{O}[\omega_1(f; 1/\sqrt{n})]$  in Theorem 8 might be improved to the order  $\mathcal{O}[\omega_1(f; \ln(n)/n)]$ .

3) It is interesting to note that the same estimate could be obtained for the truncated version of the Favard-Szász-Mirakjan operator

$$F_n^{(M)}(f)(x) = \frac{\sum_{k=0}^n \frac{(nx)^k}{k!} f\left(\frac{k}{n}\right)}{\sum_{k=0}^n \frac{(nx)^k}{k!}}, x \in [0, \infty), n \in \mathbb{N}.$$

## 5 Applications

First it is worth noting that from computational point of view, the nonlinear operators in the previous sections require a less amount of computation than their linear counterpart, since the computation of "max" is faster than that of the "sum". Also, if we take into account that the maximal term in the denominator is found explicitly in the proofs of the proposed theorems, then this does not require an extra loop for computation.

In what follows, we illustrate by two concrete examples another possible usefulness of the proposed nonlinear operators.

Thus, let us consider the functions  $f_1, f_2 : [0, 1] \rightarrow \mathbb{R}_+$

$$f_1(x) = 2 + \sin \frac{1}{x + 0.1}$$

and

$$f_2(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 0.4 \\ 10x - 3 & \text{if } 0.4 < x \leq 0.5 \\ 2 & \text{if } 0.5 < x \leq 1 \end{cases}.$$

In Figs. 1 and 2 we compare the linear and nonlinear Bernstein and Favard-Szász-Mirakjan operators approximating the function  $f_1$ .

We observe that the nonlinear operators of either Bernstein or Favard-Szász-Mirakjan type are able to better follow the peaks of the original function.

In Figs. 3 and 4 we compare the linear and nonlinear Bernstein and Favard-Szász-Mirakjan operators associated with the function  $f_2$ .

**Remark 10** *It is clear that in contrast to their linear counterparts, due to the "max" operator, these nonlinear operators in general do not preserve the smoothness of the function  $f(x)$ , that is, even if  $f$  has continuous derivative,  $B_n^{(M)}(f)(x)$  and  $F_n^{(M)}(f)(x)$  fail to be differentiable at some points.*

*However, as the graphs in the case of the function  $f_2(x)$  show, for the approximation of functions which are not differentiable at some points, these operators seem to be more suitable than their linear counterparts.*



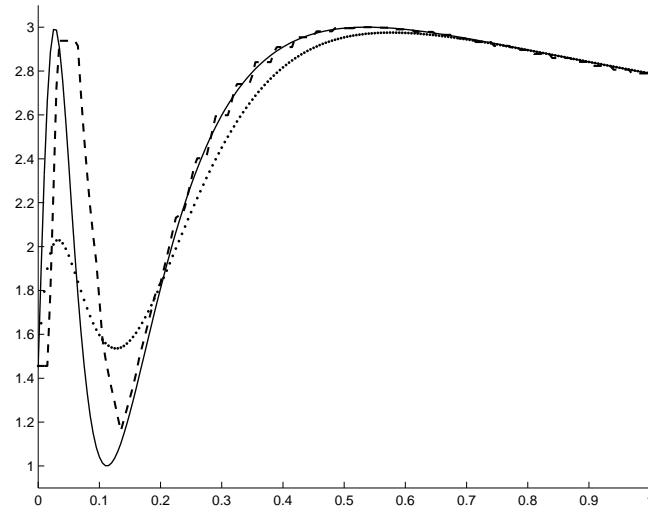


Figure 1: Comparison of Bernstein operators. Solid line:  $f_1$ , dotted line: linear case, dashed line: nonlinear case.

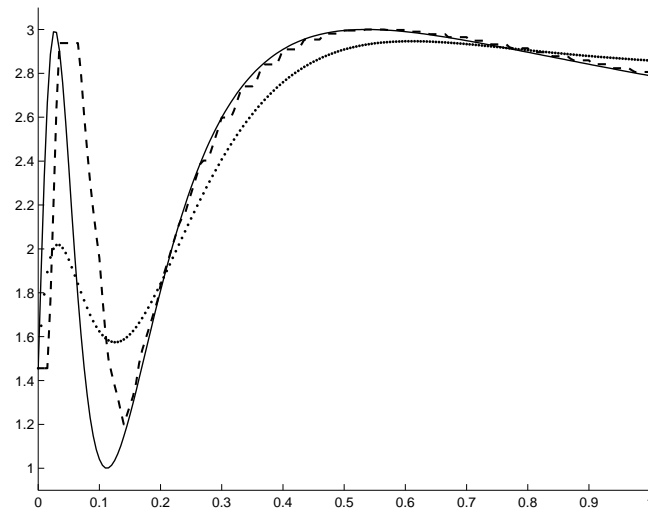


Figure 2: Comparison of Favard-Szász-Mirakjan operators. Solid line:  $f_1$ , dotted line: linear case, dashed line: nonlinear case.

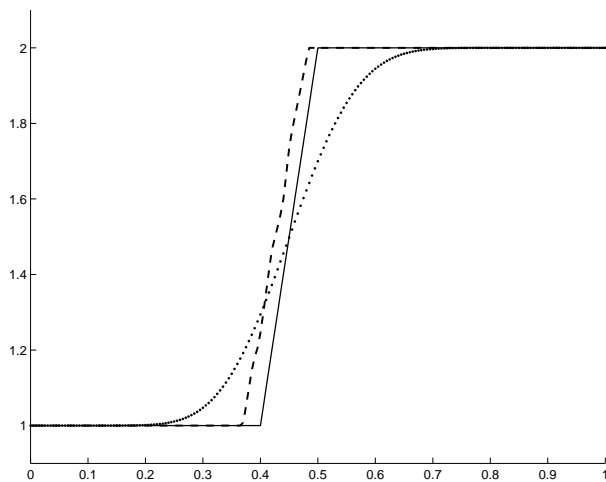


Figure 3: Comparison of Bernstein operators. Solid line:  $f_2$ , dotted line: linear case, dashed line: nonlinear case.

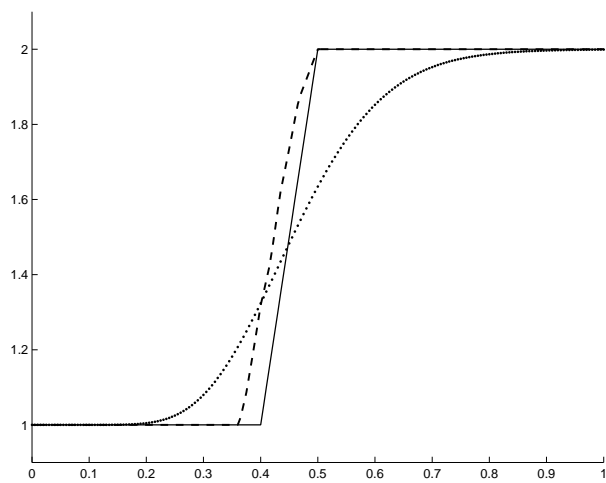


Figure 4: Comparison of Favard-Szász-Mirakjan operators. Solid line:  $f_2$ , dotted line: linear case, dashed line: nonlinear case.

## References

- [1] B. Bede, H. Nobuhara, J. Fodor, K. Hirota, Max-product Shepard approximation operators, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **10**(2006), 494–497.
- [2] B. Bede, H. Nobuhara, M. Daňková, A. Di Nola, Approximation by pseudo-linear operators, *Fuzzy Sets and Systems* **159**(2008) 804 – 820.
- [3] R.A. DeVore, V.N. Temlyakov, Nonlinear approximation in finite-dimensional spaces, *J. of Complexity*, **13**(1997), 489–508.
- [4] S.G. Gal, Shape-Preserving Approximation by Real and Complex Polynomials, Birkhäuser, Boston-Basel-Berlin, 2008.
- [5] A. Hofinger, Nonlinear function approximation : Computing smooth solutions with an adaptive greedy algorithm, *Journal of Approximation Theory*, **143**(2006) 159–175.

**$L$ -APPROXIMATION TO NON-PERIODIC FUNCTIONS**

MICHAEL I. GANZBURG

Department of Mathematics, Hampton University, Hampton, Virginia 23668,  
 E-mail: michael.ganzburg@hamptonu.edu

**ABSTRACT.** Some problems of finding exact values of the errors  $A_\sigma(f)_{L(\mathbb{R})}$  of best approximation by entire functions of exponential type in integral metrics are discussed. In particular, we prove generalized Markov- and Nagy-type theorems and apply them to find  $A_\sigma(f)_{L(\mathbb{R})}$  for some even and odd functions, such as  $|x|^\lambda$ ,  $\operatorname{sgn}(x)|x|^\lambda$ , and  $x^n \log|x|$ .

**KEY WORDS:** Best approximation, entire functions of exponential type.

## 1. INTRODUCTION

In this paper we discuss some problems of finding exact values of the errors of best approximation by entire functions of exponential type in the integral metric.

Let  $L(\mathbb{R})$  be the Banach space of all functions  $f$  with the finite norm  $\|f\|_{L(\mathbb{R})} := \int_{\mathbb{R}} |f(x)| dx$  and let  $B_\sigma$  be the class of all entire functions of exponential type  $\sigma > 0$ . We define the error of best approximation by functions from  $B_\sigma$  to a locally integrable function  $f : \mathbb{R} \rightarrow \mathbb{R}$  by

$$A_\sigma(f)_{L(\mathbb{R})} := \inf_{g_\sigma \in B_\sigma} \|f - g_\sigma\|_{L(\mathbb{R})}.$$

The Fourier transform of a function or a tempered distribution  $f$  is denoted by  $\mathcal{F}(f)$ ; similarly, the cos-transform of an even function or a tempered distribution  $f$  is denoted by  $\mathcal{F}_c(f)$  and the sin-transform of an odd function or a tempered distribution  $f$  is denoted by  $\mathcal{F}_s(f)$ . In particular for  $f \in L(\mathbb{R})$ ,

$$\begin{aligned} \mathcal{F}_c(f)(t) &:= \int_{\mathbb{R}} f(x) \cos tx \, dx, & \mathcal{F}_s(f)(t) &:= \int_{\mathbb{R}} f(x) \sin tx \, dx, \\ \mathcal{F}(f)(t) &:= \int_{\mathbb{R}} f(x) \exp(itx) \, dx. \end{aligned}$$

Study of the problem of finding  $A_\sigma(f)_{L(\mathbb{R})}$  for some continuous functions  $f$  was initiated by M. Krein [13] (see also [1, Sec. 87]) in 1938 who proved the following result:

**Theorem 1.1.** *Let a continuous function  $f$  satisfy the inequality  $|f(x)| \leq C(1 + x^2)^{-1}$ ,  $x \in \mathbb{R}$ , and let there exist a number  $\alpha \in [0, \pi/\sigma]$  and an entire function  $g_\sigma \in B_\sigma \cap L(\mathbb{R})$  such that the following  $M_\sigma$ -condition holds: the product  $\sin[\sigma(x - \alpha)](f(x) - g_\sigma(x))$  does not change its sign for all  $x \in \mathbb{R}$ . Then  $g_\sigma$  is an entire*

MICHAEL I. GANZBURG

function of best approximation to  $f$  in  $L(\mathbb{R})$  and

$$\begin{aligned} A_\sigma(f)_{L(\mathbb{R})} &= \left| \int_{\mathbb{R}} f(x) \operatorname{sgn} \sin[\sigma(x - \alpha)] dx \right| \\ &= \frac{4}{\pi} \left| \sum_{k=0}^{\infty} \frac{1}{2k+1} \operatorname{Im}(e^{-i\alpha(2k+1)\sigma} \mathcal{F}(f)((2k+1)\sigma)) \right|. \end{aligned}$$

Note that the  $M_\sigma$ -condition of Theorem 1.1 means that  $g_\sigma$  interpolates  $f$  at equidistant points of  $\mathbb{R}$  and the difference  $f - g_\sigma$  changes its sign at these and only these points. We also remark that  $\alpha$  cannot be chosen arbitrarily in Theorem 1.1 since  $\alpha$  is a solution of the equation

$$\sum_{k=-\infty}^{\infty} (-1)^k f(\alpha + k\pi/\sigma) = 0.$$

In 1939 Nagy [16] (see also [1, Sec. 88]) found classes of odd and even functions on  $\mathbb{R}$  that satisfy the  $M_\sigma$ -condition.

**Theorem 1.2.** (a) Let  $f$  be an even function from  $L(\mathbb{R})$  such that its cos-transform  $\mathcal{F}_c(f)$  is twice differentiable on  $\mathbb{R}$ , thrice differentiable on  $[\sigma, \infty)$ , and for  $t > \sigma$ ,

$$\mathcal{F}_c(f)(t) > 0, \quad \frac{d\mathcal{F}_c(f)(t)}{dt} \leq 0, \quad \frac{d^2\mathcal{F}_c(f)(t)}{dt^2} \geq 0, \quad \frac{d^3\mathcal{F}_c(f)(t)}{dt^3} \geq 0.$$

Then  $f$  satisfies  $M_\sigma$ -condition with  $\alpha = \pi/(2\sigma)$  and

$$A_\sigma(f)_{L(\mathbb{R})} = \frac{4}{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{\mathcal{F}_c(f)((2k+1)\sigma)}{2k+1}. \quad (1.1)$$

(b) Let  $f$  be an odd function from  $L(\mathbb{R})$  such that its sin-transform  $\mathcal{F}_s(f)$  is twice differentiable on  $\mathbb{R}$  and for  $t > \sigma$ ,

$$\mathcal{F}_s(f)(t) > 0, \quad \frac{d\mathcal{F}_s(f)(t)}{dt} \leq 0, \quad \frac{d^2\mathcal{F}_s(f)(t)}{dt^2} \geq 0.$$

Then  $f$  satisfies  $M_\sigma$ -condition with  $\alpha = 0$  and

$$A_\sigma(f)_{L(\mathbb{R})} = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{\mathcal{F}_s(f)((2k+1)\sigma)}{2k+1}. \quad (1.2)$$

Theorems 1.1 and 1.2 have been used in approximation theory for finding exact constants of approximation on convolution classes (see [1, 20, 12]).

Since 1985 Vaaler and his students have published several papers on best approximation to some locally integrable functions. Their research has been influenced by applications of these results to some problems of number theory. In particular, Vaaler [21, Th. 4] established the relation

$$A_\sigma(\operatorname{sgn} x)_{L(\mathbb{R})} = \pi/\sigma. \quad (1.3)$$

Littmann [14, Th. 6.2] generalized this result by proving the relation

$$A_\sigma(\operatorname{sgn}(x) x^n)_{L(\mathbb{R})} = \frac{8n!}{\pi\sigma^{n+1}} \sum_{k=0}^{\infty} \frac{(-1)^{kn}}{(2k+1)^{n+2}}, \quad n = 0, 1, \dots \quad (1.4)$$

Since  $\sum_{k=0}^{\infty} (2k+1)^{-2} = \pi^2/8$ , (1.3) follows from (1.4).

In recent paper [4, Sec. 1] Carneiro and Vaaler established the following results:

$$A_\sigma(|x|^\lambda)_{L(\mathbb{R})} = \frac{8\Gamma(\lambda+1)\sin(\pi\lambda/2)}{\pi^{\lambda+2}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^{\lambda+2}}, \quad |\lambda| \leq 1, \quad (1.5)$$

$$A_\sigma(\log|x|)_{L(\mathbb{R})} = \frac{4}{\sigma} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2}. \quad (1.6)$$

Note that Theorems 1.1 and 1.2 cannot be applied to these functions since all of them do not belong to  $L(R)$ .

In this paper we extend Theorems 1.1 and 1.2 to locally integrable functions on  $\mathbb{R}$  and show that relations (1.3) through (1.6) are easy corollaries of these results. Other examples are discussed as well.

## 2. KREIN- AND NAGY-TYPE THEOREMS FOR LOCALLY INTEGRABLE FUNCTIONS

We first discuss a more general version of Theorem 1.1.

**Theorem 2.1.** *Let  $f$  be a locally integrable function on  $\mathbb{R}$  and let there exist a number  $\alpha \in [0, \pi/\sigma]$  and an entire function  $g_\sigma \in B_\sigma$  such that  $f - g_\sigma \in L(\mathbb{R})$  and the following  $M_\sigma^*$ -condition holds: the product  $\sin[\sigma(x - \alpha)](f(x) - g_\sigma(x))$  does not change its sign for a. a.  $x \in \mathbb{R}$ . Then  $g_\sigma$  is an entire function of best approximation to  $f$  in  $L(\mathbb{R})$  and*

$$A_\sigma(f)_{L(\mathbb{R})} = \left| \int_{\mathbb{R}} (f(x) - g_\sigma(x)) \operatorname{sgn} \sin[\sigma(x - \alpha)] dx \right|.$$

*Proof.* If  $g \in B_\sigma \cap L(\mathbb{R})$ , then  $\mathcal{F}(g)(t)$  is continuous on  $\mathbb{R}$ . Therefore, by the Paley-Wiener theorem [19, Th. 7.2.1],  $\mathcal{F}(g)(t) = 0$  for  $|t| \geq \sigma$ . Next using properties of Fourier sums of functions of bounded variation over  $(0, 2\pi)$  (see [22, Th. 2.8.1] and [1, Sec. 53]), we have

$$\frac{4}{\pi} \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{\sin[(2k+1)\sigma(x - \alpha)]}{2k+1} = \operatorname{sgn} \sin[\sigma(x - \alpha)], \quad x \in \mathbb{R},$$

and

$$\sup_{x \in \mathbb{R}} \sup_{n \geq 0} \left| \frac{4}{\pi} \sum_{k=0}^n \frac{\sin[(2k+1)\sigma(x - \alpha)]}{2k+1} \right| < \infty.$$

Then by the Lebesgue dominated convergence theorem, for any  $g \in L(\mathbb{R})$  we have

$$\begin{aligned} \int_{\mathbb{R}} |f(x) - g_\sigma(x) - g(x)| dx &\geq \left| \int_{\mathbb{R}} (f(x) - g_\sigma(x) - g(x)) \operatorname{sgn} \sin[\sigma(x - \alpha)] dx \right| \\ &= \left| \lim_{n \rightarrow \infty} \frac{4}{\pi} \int_{\mathbb{R}} (f(x) - g_\sigma(x) - g(x)) \sum_{k=0}^n \frac{\sin[(2k+1)\sigma(x - \alpha)]}{2k+1} dx \right| \\ &= \left| \lim_{n \rightarrow \infty} \frac{4}{\pi} \int_{\mathbb{R}} (f(x) - g_\sigma(x)) \sum_{k=0}^n \frac{\sin[(2k+1)\sigma(x - \alpha)]}{2k+1} dx \right| \\ &= \left| \int_{\mathbb{R}} (f(x) - g_\sigma(x)) \operatorname{sgn} \sin[\sigma(x - \alpha)] dx \right| = \int_{\mathbb{R}} |f(x) - g_\sigma(x)| dx. \end{aligned}$$

This proves Theorem 2.1. □

**Remark 2.1.** Note that Theorem 2.1 strengthens Theorem 1.1 even in the case  $f \in L(\mathbb{R})$  since we replace a condition  $|f(x)| \leq C(1+x^2)^{-1}$  by a weaker condition  $f \in L(\mathbb{R})$ . In addition, note that the proof of Theorem 2.1 is similar to that of Theorem 1.1 (cf. [1, Sec. 87]). We remark also that a different version of Theorem 2.1 was proved earlier by the author [5].

Next we prove a Nagy-type theorem for locally integrable functions.

**Theorem 2.2.** (a) Let  $f$  be an even locally integrable function on  $\mathbb{R}$ , which is a tempered distribution, and let for some  $\sigma_0 \geq 0$  the restriction to  $(-\infty, -\sigma_0) \cup (\sigma_0, \infty)$  of the cos-transform  $\mathcal{F}_c(f)$  of the tempered distribution  $f$  be a thrice differentiable function. If  $\mathcal{F}_c(f)$  satisfies the following conditions for  $t > \sigma_0$ :

$$\mathcal{F}_c(f)(t) > 0, \quad \frac{d\mathcal{F}_c(f)(t)}{dt} \leq 0, \quad \frac{d^2\mathcal{F}_c(f)(t)}{dt^2} \geq 0, \quad \frac{d^3\mathcal{F}_c(f)(t)}{dt^3} \geq 0, \quad (2.1)$$

and

$$\lim_{t \rightarrow \infty} \mathcal{F}_c(f)(t) = 0, \quad (2.2)$$

then  $f$  satisfies  $M_\sigma^*$ -condition with  $\alpha = \pi/(2\sigma)$  and  $\sigma > \sigma_0$ . In addition,

$$A_\sigma(f)_{L(\mathbb{R})} = \frac{4}{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{\mathcal{F}_c(f)((2k+1)\sigma)}{2k+1}.$$

(b) Let  $f$  be an odd locally integrable function on  $\mathbb{R}$ , which is a tempered distribution, and let for some  $\sigma_0 \geq 0$  the restriction to  $(-\infty, -\sigma_0) \cup (\sigma_0, \infty)$  of the sin-transform  $\mathcal{F}_s(f)$  of the tempered distribution  $f$  be a twice differentiable function. If  $\mathcal{F}_s(f)$  satisfies the following conditions for  $t > \sigma_0$ :

$$\mathcal{F}_s(f)(t) > 0, \quad \frac{d\mathcal{F}_s(f)(t)}{dt} \leq 0, \quad \frac{d^2\mathcal{F}_s(f)(t)}{dt^2} \geq 0, \quad (2.3)$$

and

$$\lim_{t \rightarrow \infty} \mathcal{F}_s(f)(t) = 0, \quad (2.4)$$

then  $f$  satisfies  $M_\sigma^*$ -condition with  $\alpha = 0$  and  $\sigma > \sigma_0$ . In addition,

$$A_\sigma(f)_{L(\mathbb{R})} = \frac{4}{\pi} \sum_{k=0}^{\infty} \frac{\mathcal{F}_s(f)((2k+1)\sigma)}{2k+1}.$$

*Proof.* (a) Let  $\sigma > \sigma_0$ . We first find a function  $h_\sigma \in B_\sigma$  such that  $f - h_\sigma \in L(R)$ . Setting  $\tau := (\sigma_0 + \sigma)/2$ , we extend  $\mathcal{F}_c(f)(t)$  from  $(-\infty, -\tau] \cup [\tau, \infty)$  to  $\mathbb{R}$  by the formula

$$F(t) := \begin{cases} \mathcal{F}_c(f)(t), & |t| \geq \tau \\ P_4(t), & |t| < \tau, \end{cases} \quad (2.5)$$

where

$$P_4(t) := \frac{\tau \mathcal{F}^{(2)} - \mathcal{F}^{(1)}}{8\tau^3} t^4 + \frac{-\tau \mathcal{F}^{(2)} + 3\mathcal{F}^{(1)}}{4\tau} t^2 + \frac{\tau^2 \mathcal{F}^{(2)} - 5\tau \mathcal{F}^{(1)} + 8\mathcal{F}^{(0)}}{8}$$

is the Hermite polynomial satisfying the relations

$$P_4^{(s)}(\pm\tau) = \mathcal{F}^{(s)}(\pm\tau) := \left. \frac{d^{(s)}[\mathcal{F}_c(f)(t)]}{dt^s} \right|_{t=\pm\tau}, \quad s = 0, 1, 2.$$

## L-APPROXIMATION TO NON-PERIODIC FUNCTIONS

Then  $F$  is an even and bounded function on  $\mathbb{R}$ . Moreover, it is a twice differentiable function on  $\mathbb{R}$  and a thrice differentiable function on  $[\tau, \infty)$ , which satisfies the conditions

$$F(t) > 0, \quad F'(t) \leq 0, \quad F''(t) \geq 0, \quad F'''(t) \leq 0, \quad t > \tau, \quad (2.6)$$

and

$$\lim_{t \rightarrow \infty} F(t) = 0. \quad (2.7)$$

Next, it follows from (2.5) and (2.6) that  $F' \in L(\mathbb{R})$  and  $F'' \in L(\mathbb{R})$ . Hence integrating by parts and taking account of (2.7), we have for  $x \neq 0$

$$\begin{aligned} \mathcal{F}_c(F)(x) &= \lim_{A \rightarrow \infty} \int_{-A}^A F(t) \cos xt \, dt \\ &= \lim_{A \rightarrow \infty} \left( \frac{F(t) \sin xt}{x} \Big|_{-A}^A - \frac{1}{x} \int_{-A}^A F'(t) \sin xt \, dt \right) \\ &= -\frac{1}{x} \int_{\mathbb{R}} F'(t) \sin xt \, dt = \frac{1}{x^2} \int_{\mathbb{R}} F''(t) \cos xt \, dt. \end{aligned}$$

Hence  $\mathcal{F}_c(F)(x)$  exists for every  $x \neq 0$  and

$$|\mathcal{F}_c(F)(x)| \leq Cx^{-2}, \quad x \neq 0. \quad (2.8)$$

Further setting  $\varphi(x) := (2\pi)^{-1} \mathcal{F}_c(F)(x)$ , we shall show that  $h_\sigma := f - \varphi$  belongs to  $B_\sigma$ . Indeed,  $\varphi$  is a tempered distribution since it is the cos-transform of a bounded continuous function  $(2\pi)^{-1} F(t)$  on  $\mathbb{R}$ . Hence

$$\mathcal{F}_c(\varphi) = F. \quad (2.9)$$

Then  $h_\sigma$  is an even tempered distribution, and its cos-transform  $\mathcal{F}_c(h_\sigma)$  is defined as the functional  $(h_\sigma, \mathcal{F}_c(\psi))$ , where  $\psi$  is an even rapidly decreasing function from the Schwartz class  $S$ . Therefore, if  $\psi \in S$  and  $\psi = 0$  on  $[-\tau_1, \tau_1]$ , where  $\tau_1 \in (\tau, \sigma)$ , then by (2.5) and (2.9),

$$(h_\sigma, \mathcal{F}_c(\psi)) = (f, \mathcal{F}_c(\psi)) - (\varphi, \mathcal{F}_c(\psi)) = \int_{|t| \geq \tau_1} (\mathcal{F}_c(f)(t) - F(t)) \psi(t) \, dt = 0.$$

Thus the support of  $\mathcal{F}_c(h_\sigma)$  is a subset of  $[-\sigma, \sigma]$ . Using now the generalized Paley-Wiener theorem [19, Th. 7.2.3], we arrive at  $h_\sigma \in B_\sigma$ . Next,

$$f(x) - h_\sigma(x) = (2\pi)^{-1} \mathcal{F}_c(F)(x), \quad (2.10)$$

so (2.8) and (2.10) imply

$$\begin{aligned} \int_{\mathbb{R}} |f(x) - h_\sigma(x)| \, dx &\leq \int_{|x| \leq 1} |f(x)| \, dx + \int_{|x| \leq 1} |h_\sigma(x)| \, dx \\ &\quad + \int_{|x| > 1} |f(x) - h_\sigma(x)| \, dx < \infty. \end{aligned}$$

Thus  $\varphi = f - h_\sigma \in L(\mathbb{R})$ . Moreover, since  $\varphi \in L(\mathbb{R})$ , identity (2.9) holds not only in the distributional sense but also as a formula for the cos-transform of an integrable function. Hence taking account of (2.5), (2.6), and (2.9), we conclude that  $\varphi$  satisfies all the conditions of Theorem 1.2(a). Therefore,  $\varphi$  satisfies  $M_\sigma$ -condition for  $\alpha = \pi/(2\sigma)$ , that is, there exists  $G_\sigma \in B_\sigma \cap L(\mathbb{R})$  such that the function  $\cos \sigma x [f(x) - h_\sigma(x) - G_\sigma(x)]$  does not change its sign on  $\mathbb{R}$ . Therefore,



MICHAEL I. GANZBURG

we conclude that  $f$  satisfies  $M_\sigma^*$ -condition for  $\alpha = \pi/(2\sigma)$  and  $g_\sigma := h_\sigma + G_\sigma$ . Moreover, by (1.1), (2.5), and (2.9),

$$\begin{aligned} A_\sigma(f)_{L(\mathbb{R})} = A_\sigma(f - h_\sigma)_{L(\mathbb{R})} &= \frac{4}{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{\mathcal{F}_c(f - h_\sigma)((2k+1)\sigma)}{2k+1} \\ &= \frac{4}{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{\mathcal{F}_c(f)((2k+1)\sigma)}{2k+1}. \end{aligned}$$

This proves statement (a) of Theorem 2.2.

(b) The proof of this statement is similar to that of Theorem 2.2(a) if we replace  $\mathcal{F}_c(f)$  by  $\mathcal{F}_s(f)$  and  $P_4(t)$  by the polynomial

$$\begin{aligned} P_5(t) &:= \frac{\tau^2 \mathcal{F}^{(2)} - 3\tau \mathcal{F}^{(1)} + 3\mathcal{F}^{(0)}}{8\tau^5} t^5 + \frac{-\tau^2 \mathcal{F}^{(2)} + 5\tau \mathcal{F}^{(1)} - 5\mathcal{F}^{(0)}}{4\tau^3} t^3 \\ &\quad + \frac{\tau^2 \mathcal{F}^{(2)} - 7\tau \mathcal{F}^{(1)} + 15\mathcal{F}^{(0)}}{8\tau} t. \end{aligned}$$

Therefore, Theorem 2.2 is established.  $\square$

**Remark 2.2.** Note that Theorem 2.2 strengthens Theorem 1.2 even in the case  $f \in L(\mathbb{R})$  since we drop the conditions that  $\mathcal{F}_c(f)$  and  $\mathcal{F}_s(f)$  should be twice differentiable on  $\mathbb{R}$ . We remark also that a special case of Theorem 2.2 for  $\sigma_0 = 0$  was proved earlier by the author [5].

### 3. EXAMPLES

**Example 3.1.**  $f_{n,\text{sgn}}(x) := \text{sgn}(x) x^n$ ,  $n = 0, 1, 2, \dots$

Then the Fourier transforms of  $f_{n,\text{sgn}}$  are given by the formulas [3, Sec. 10.1, #6]

$$\begin{aligned} \mathcal{F}_c(f_{n,\text{sgn}})(t) &= 2n!(-1)^{(n+1)/2} t^{-(n+1)}, \quad t > 0, \quad n = 1, 3, 5, \dots, \\ \mathcal{F}_s(f_{n,\text{sgn}})(t) &= 2n!(-1)^{n/2} t^{-(n+1)}, \quad t > 0, \quad n = 0, 2, 4, \dots \end{aligned} \quad (3.1)$$

Since for odd  $n \geq 1$ ,  $(-1)^{(n+1)/2} f_{n,\text{sgn}}$  satisfies all the conditions of Theorem 2.2(a) and for even  $n \geq 0$ ,  $(-1)^{n/2} f_{n,\text{sgn}}$  satisfies all the conditions of Theorem 2.2(b), we arrive at (1.4).

**Example 3.2.**  $f_\lambda(x) := |x|^\lambda$ ,  $\lambda \in (-1, \infty)$ ,  $\lambda \neq 0, 2, 4, \dots$

Then by [3, Sec. 10.1, #11],

$$\mathcal{F}_c(f_\lambda)(t) = -2 \sin(\lambda\pi/2) \Gamma(\lambda+1) t^{-(\lambda+1)}, \quad t > 0.$$

Since  $-\sin(\lambda\pi/2) f_\lambda$  satisfies the conditions of Theorem 2.2(a), we obtain

$$A_\sigma(f_\lambda)_{L(\mathbb{R})} = \frac{8|\sin(\lambda\pi/2)|\Gamma(\lambda+1)}{\pi\sigma^{\lambda+1}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^{\lambda+2}}. \quad (3.2)$$

**Remark 3.1.** The first direct proof of this result was given in [5]. Equality (3.2) can be also obtained as a corollary of a limit theorem for polynomial approximations in the integral metric [18] and  $L$ -approximation asymptotic results by Nikolskii [17] and Bernstein [2] (see [6, 7, 15, 10] for more details). A different proof of (3.2) for  $|\lambda| \leq 1$  was given in [4].

**Example 3.3.**  $f_{\lambda, \text{sgn}}(x) := \text{sgn}(x) |x|^\lambda$ ,  $\lambda \in (-1, \infty)$ ,  $\lambda \neq 1, 3, 5, \dots$

Then by [3, Sec. 10.1, #12],

$$\mathcal{F}_s(f_{\lambda, \text{sgn}})(t) = 2 \cos(\lambda\pi/2) \Gamma(\lambda + 1) t^{-(\lambda+1)}, \quad t > 0.$$

Since  $\cos(\lambda\pi/2)f_\lambda$  satisfies the conditions of Theorem 2.2(b), we obtain

$$A_\sigma(f_{\lambda, \text{sgn}})_{L(\mathbb{R})} = \frac{8 |\cos(\lambda\pi/2)| \Gamma(\lambda + 1)}{\pi \sigma^{\lambda+1}} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^{\lambda+2}}. \quad (3.3)$$

**Example 3.4.**  $f_{n, \log}(x) := x^n \log |x|$ ,  $n = 0, 1, 2, \dots$

Then by [3, Sec. 10.1, #s 22,23],

$$\begin{aligned} \mathcal{F}_c(f_{n, \log})(t) &= \pi (-1)^{n/2} n! t^{-(n+1)}, \quad t > \sigma_0 \geq 0, \quad n = 0, 2, 4, \dots \\ \mathcal{F}_s(f_{n, \log})(t) &= \pi (-1)^{(n+1)/2} n! t^{-(n+1)}, \quad t > \sigma_0 \geq 0, \quad n = 1, 3, 5, \dots \end{aligned}$$

Using Theorem 2.2, we get

$$\begin{aligned} A_\sigma(f_{n, \log})_{L(\mathbb{R})} &= \frac{4n!}{\sigma^{n+1}} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^{n+2}} \quad n = 0, 2, 4, \dots \\ A_\sigma(f_{n, \log})_{L(\mathbb{R})} &= \frac{4n!}{\sigma^{n+1}} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^{n+2}} \quad n = 1, 3, 5, \dots \end{aligned} \quad (3.4)$$

A different proof of (3.4) for  $n=0$  was given in [4]. A two-dimensional version of (3.4) was established in [8].

**Example 3.5.**  $f_{\beta, \text{arc}}^* := (2/\pi) \arctan(x/\beta)$ ,  $\beta > 0$ .

We first note that

$$\mathcal{F}_s(f_{\beta, \text{arc}}^*)(t) := 2t^{-1} e^{-\beta t}, \quad t > 0.$$

Indeed, using (3.1) and [11, Sec. 4.57], we have

$$\mathcal{F}_s(f_{1, \text{arc}}^*)(t) = \mathcal{F}_s(\text{sgn } x)(t) - (2/\pi) \mathcal{F}_s(\arctan x)(t) = 2/t - (2/t)(1 - e^{-t}) = 2t^{-1} e^{-t}.$$

Since  $f_{\beta, \text{arc}}^*$  satisfies the conditions of Theorem 2.2(b), we obtain

$$A_\sigma(f_{\beta, \text{arc}}^*)_{L(\mathbb{R})} = \frac{8}{\pi \sigma} \sum_{k=0}^{\infty} \frac{1}{(2k+1)^2 e^{\beta \sigma (2k+1)}}.$$

Note that for all  $x \in \mathbb{R}$ ,  $\lim_{\beta \rightarrow 0+} f_{\beta, \text{arc}}^*(x) = \text{sgn } x$  and

$$\lim_{\beta \rightarrow 0+} A_\sigma(f_{\beta, \text{arc}}^*)_{L(\mathbb{R})} = A_\sigma(\text{sgn } x)_{L(\mathbb{R})} = \pi/\sigma.$$

**Remark 3.2.** Similarly to Examples 3.1-3.5, we can find  $A_\sigma(f)_{L(\mathbb{R})}$  for functions such as  $\text{sgn}(x) x^n \log |x|$ ,  $|x|^\lambda \log |x|$ ,  $\text{sgn}(x) |x|^\lambda \log |x|$ ,  $\lambda \neq 0, 1, 2, \dots$ , as well as more general functions of the form  $|x|^\lambda \log^k |x|$  and  $\text{sgn } |x|^\lambda \log^k |x|$ . Note that some of these results were recently discussed in [9, pp. 94, 95].

MICHAEL I. GANZBURG

## REFERENCES

- [1] N.I. Akhiezer, *Lectures on the Theory of Approximation*, (2nd ed.), Nauka, Moscow, 1965. [Russian]
- [2] S.N. Bernstein, On the best approximation of  $|x - c|^p$ , in *Collected Works*, Vol II, Akad. Nauk SSSR, Moscow, 1954, pp.273–280. [Russian]
- [3] Yu.A. Brychkov and A.P. Prudnikov, *Integral Transforms of Generalized Functions*, Gordon and Breach Science Publishers, New York, 1989.
- [4] E. Carneiro and J.D. Vaaler, Some extremal functions in Fourier analysis, III, arXiv:0809.4053v1 [math.CA] 23 Sep 2008.
- [5] M.I. Ganzburg, Criteria for best approximation of locally integrable functions in  $L(\mathbb{R})$ , in *Studies in Current Problems of Summation and Approximation of Functions and their Applications*, Dnepropetrovsk Gos. University, Dnepropetrovsk, 1983, pp. 11–16. [Russian]
- [6] M.I. Ganzburg, Limit theorems and best constants of approximation theory, in *Handbook on Analytic-Computational Methods in Applied Mathematics* (G.A. Anastassiou, ed.), CRC Press, Boca Raton, FL, 2000, pp.507–569.
- [7] M.I. Ganzburg, The Bernstein constant and polynomial interpolation at the Chebyshev nodes, *J. Approx. Theory*, 119,193–213(2002).
- [8] M.I. Ganzburg, Best constants of harmonic approximation on classes associated with the Laplace operator, *J. Approx. Theory*, 150,199–213(2008).
- [9] M.I. Ganzburg, *Limit Theorems of Polynomial Approximation with Exponential Weights*, Memoirs of AMS, 897, American Mathematical Society, Providence, RI, 2008.
- [10] M.I. Ganzburg and D.S. Lubinsky, Best approximating entire functions to  $|x|^\alpha$  in  $L_2$ , *Contemporary Math.*, 455,93–107(2008).
- [11] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series and Products*, Academic Press, San Diego, 1980.
- [12] N.P. Korneichuk, *Exact Constants in Approximation Theory*, Cambridge University Press, Cambridge, 1991.
- [13] M.G. Krein, On the best approximation of continuous differentiable functions on the whole real axis, *Dokl. Akad. Nauk SSSR*, 18,615–624(1938). [Russian]
- [14] F. Littman, Entire approximations to the truncated powers, *Constr. Approx.*, 22,273–295(2005).
- [15] D.S. Lubinsky, Series representations for best approximating entire functions of exponential type, in *Proceedings of the International Conference on the Interactions between Wavelets and Splines*, Athens, GA, Nashboro Press, Brentwood, TN, 2006, pp. 356–364.
- [16] B. Sz.-Nagy, Über gewisse Extremalfragen bei transformierten trigonometrischen Entwicklungen. II. Nichtperiodischer Fall, *Berichte Acad. d. Wiss., Leipzig*, 91.
- [17] S.M. Nikolskii, On the best mean approximation by polynomials of the functions  $|x - c|^s$ , *Izvestia Akad. Nauk SSSR*, 11,139–180(1947). [Russian]
- [18] R.A. Ratsin, S. N. Bernstein limit theorem for the best approximation in the mean and some of its applications, *Izv. Vysch. Uchebn. Zaved. Mat.*, 12,81–86(1968).
- [19] R.S. Strichartz, *A Guide to Distribution Theory and Fourier Transforms*, CRC Press, Boca Raton, FL, 1994.
- [20] A.F. Timan, *Theory of Approximation of Functions of a Real Variable*, MacMillan, New York, 1963.
- [21] J. Vaaler, Some extremal functions in Fourier analysis, *Bull. Amer. Math. Soc.*, 12,183–216(1985).
- [22] A. Zygmund, *Trigonometric Series* (2nd ed.), Vol. I, Cambridge University Press, Cambridge, 1959.

## Inequalities for Self-Reciprocal Polynomials and Uniformly Almost Periodic Functions

N. K. Govil

Department of Mathematics and Statistics

Auburn University

Auburn, AL 36849

U. S. A.

E-mail: govilnk@auburn.edu

Q. M. Tariq

Department of Mathematics & Computer Science

Virginia State University

Petersburg, VA 23806

U. S. A.

E-mail: tqazi@vsu.edu

### Abstract

If  $p(z)$  is a polynomial of degree  $n$  and  $p'(z)$  its derivative, then it is well known that

$$\max_{|z|=1} |p'(z)| \leq n \max_{|z|=1} |p(z)|.$$

Also, for an entire function  $f(z)$  of exponential type  $\tau$ , it was proved by S. N. Bernstein that

$$\sup_{-\infty < x < \infty} |f'(x)| \leq \tau \sup_{-\infty < x < \infty} |f(x)|.$$

Both the above inequalities are sharp, and have been the starting point of a considerable literature in Approximation Theory. In this paper we will discuss some of the research centered around above inequalities for some classes of polynomials, including self-reciprocal polynomials, and the extension of some of these results for entire functions of exponential type, and for uniformly almost periodic functions.

**Key Words:** Self-reciprocal polynomials, Entire functions of exponential type, Uniformly almost periodic functions.

## 1 Introduction

Some years after the chemist Mendeleieff invented the periodic table of the elements he made a study of the specific gravity of a solution as a function of the percentage of the dissolved substance [38]. This function is of some practical importance: for example it is used in testing beer and wine for alcoholic content, and in testing the cooling system of an automobile for concentration of anti-freeze; but present-day physical chemists do not seem to find it as interesting as Mendeleieff did. Nevertheless Mendeleieff's study led to mathematical problems of great interest, some of which are even today inspiring research in Mathematics.

An example of the kind of curve that Mendeleieff obtained was specific gravity of a solution in terms of the percentage of alcohol. He noticed that such curves could be closely approximated by successions of quadratic arcs and he wanted to know whether the corners where the arcs joined were really there, or just caused by errors of measurement. In mathematical terms, this amounts to considering a quadratic polynomial  $P(x) = px^2 + qx + r$  with  $|P(x)| \leq 1$  for  $-1 \leq x \leq 1$ , and estimating how large can  $|P'(x)|$  be on  $-1 \leq x \leq 1$  (for details, how the Mendeleieff's problem in Chemistry amounts to this mathematical problem in polynomials, see Boas [7]). Surprisingly, Mendeleieff himself was able to solve this mathematical problem and prove that  $|P'(x)| \leq 4$ ; and this inequality is best possible, since when  $P(x) = 1 - 2x^2$  we have  $|P(x)| \leq 1$  for  $-1 \leq x \leq 1$  and  $|P'(\pm 1)| = 4$ . By using this result Mendeleieff was able to convince himself that the corners in his curve were genuine; and he was right, since his measurements were quite accurate, as they agree with modern tables to three or more significant figures.

Mendeleieff told his result to a Russian mathematician A. A. Markoff, who investigated the corresponding problem in a more general setup, that is, for polynomials of arbitrary degree  $n$ . He [37] proved the following result which is known as Markoff's Theorem.

**THEOREM 1.1** *If  $p(x) = \sum_{\nu=0}^n a_{\nu}x^{\nu}$  is a real polynomial of degree  $n$  and  $|p(x)| \leq 1$  on  $[-1, 1]$  then*

$$|p'(x)| \leq n^2 \text{ for } -1 \leq x \leq 1. \quad (1.1)$$

*The inequality is best possible and is attained at only  $x = \pm 1$  only when  $p(x) = \pm T_n(x)$ , where  $T_n(x)$  (the so called Tchebycheff polynomial of the first kind) is  $\cos(n \cos^{-1} x)$  (which actually is a polynomial, since  $\cos n\theta$  is a polynomial in  $\cos \theta$ ). In fact  $T_n(x) = \cos(n \cos^{-1} x) = 2^{n-1} \prod_{\nu=1}^n \{x - \cos((\nu - \frac{1}{2})\pi/n)\}$ .*

N.K. Govil, et al.,

It was several years later, around 1926, when a Russian mathematician Serge Bernstein needed the analogue of Theorem 1.1 for the unit disc in the complex plane instead of the interval  $[-1, 1]$ . He wanted to know if  $p(z)$  is a polynomial of degree at most  $n$  (by a polynomial of degree at most  $n$  we mean an expression of the form  $\sum_{\nu=0}^n a_{\nu} z^{\nu}$ ,  $a_{\nu}$  being complex and  $z$  a complex variable) with  $|p(z)| \leq 1$  for  $|z| \leq 1$ , then what is maximum of  $|p'(z)|$  for  $|z| \leq 1$ ?

The answer to this is given by the following which is known as Bernstein's inequality [4].

**THEOREM 1.2** *If  $p(z) = \sum_{\nu=0}^n a_{\nu} z^{\nu}$  is a polynomial of degree at most  $n$ , then*

$$\max_{|z| \leq 1} |p'(z)| \leq n \max_{|z| \leq 1} |p(z)|. \quad (1.2)$$

*The result is best possible and the equality holds for  $p(z) = \lambda z^n$ ,  $\lambda$  being a complex number.*

The above Bernstein's inequality has an analogue for trigonometric polynomials which states that if  $t(\theta) = \sum_{\nu=-n}^n a_{\nu} e^{i\nu\theta}$  is a trigonometric polynomial (possibly with complex coefficients) of degree  $n$ ,  $|t(\theta)| \leq 1$  for  $0 \leq \theta < 2\pi$  then for  $0 \leq \theta < 2\pi$ ,

$$|t'(\theta)| \leq n. \quad (1.3)$$

In (1.3) equality holds if and only if  $t(\theta) = e^{i\gamma} \cos(n\theta - \alpha)$ , where  $\gamma$  and  $\alpha$  are arbitrary real numbers.

Note that a trigonometric polynomial  $t(\theta) = \sum_{\nu=-n}^n c_{\nu} e^{i\nu\theta}$  (possibly with complex coefficients) is said to be real if  $c_{\nu} = \bar{c}_{-\nu}$ .

Inequality (1.3) is also known as Bernstein's inequality although Bernstein [4] proved (1.3) with  $2n$  in place of  $n$ . His proof was based on a variational method. Inequality (1.3) in the present form first appeared in print in a paper of Fekete [19] who attributes the proof to Fejer [17]. Bernstein [3] attributes the proof to E. Landau (see Schaeffer [51] and Fejer [18]). Alternative proofs of the inequality (1.3) have been supplied by F. Riesz [47], M. Riesz [48], de la Valee Poussin [55], Rogosinski [49] and others, and each of these methods has led to interesting extensions of the inequality (1.3).

If  $p(z) = \sum_{\nu=0}^n a_{\nu} z^{\nu}$  is a polynomial of degree at most  $n$ , then  $t(\theta) = p(e^{i\theta})$  is a trigonometric polynomial of degree  $n$  with  $|t(\theta)| \leq 1$ , real  $\theta$ , hence applying (1.3) to  $t(\theta) = p(e^{i\theta})$  one can get Theorem 1.2.

Bernstein needed the above inequalities in order to answer the following question of best approximation raised by de la Valee Poussin in the early part of last century; Is it possible to approximate every polygonal line by polynomials of degree  $n$  with an error of  $o(1/n)$  as  $n$  becomes large? (The

result that the approximation can be carried out with an error of  $0(1/n)$  was proved by de la Valee Poussin himself). This problem has played an important role in the development of the theory of approximation and was answered in the negative by S. Bernstein [4]. Several monographs and papers have been published in this area (see Boas [7], Durand [15], Govil and Mohapatra [30], Rahman and Schmeisser [42], Rassias [46], Telyakovskii [53], and Voronovskaja [56]).

The inequality (1.2) and the corresponding inequality for entire functions of exponential type (which was proved by Bernstein himself) have been the starting point of a considerable literature in approximation theory and in this paper we study some of the research centered around these inequalities. This paper consists of three sections, including Section 1, which is an introduction. In Section 2, we discuss some inequalities analogous to (1.2) for some classes of polynomials including self-reciprocal polynomials, while in Section 3 we deal with analogous inequalities for entire functions of exponential type for some of the results in Section 2, and a result for uniformly almost periodic functions.

## 2 Polynomials

We begin with Bernstein's inequality (1.2). In view of the maximum modulus principle, it states equivalently that for a polynomial  $p(z)$  of degree at most  $n$ ,

$$\max_{|z|=1} |p'(z)| \leq n \max_{|z|=1} |p(z)|. \quad (2.1)$$

The above inequality is best possible with equality holding for  $p(z) = \lambda z^n$ , for every complex number  $\lambda$ . It was conjectured by Erdős [16] that if a polynomial  $p(z)$  has no zeros in  $|z| < 1$ , then  $\max_{|z|=1} |p'(z)| \leq (n/2) \max_{|z|=1} |p(z)|$ . This conjecture was proved in the special case when  $p(z)$  has all its zeros on  $|z| = 1$  independently by Szegő and Polya (see Lax [35]). In the general case the conjecture was established for the first time by Lax [35], who proved

**THEOREM 2.1** *If  $p(z)$  is a polynomial of degree  $n$ ,  $p(z) \neq 0$  for  $|z| < 1$ , then*

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{2} \max_{|z|=1} |p(z)|. \quad (2.2)$$

*The result is best possible and the equality in (2.2) holds for any polynomial which has all its zeros on  $|z| = 1$ .*

N.K. Govil, et al.,

Simpler proofs of this result were given by de Bruijn [13] and Aziz and Mohammed [2]. For some generalizations of Theorem 2.1, see Boas [8], Gardner and Govil [23] and Rahman [40].

Professor R. P. Boas, Jr. raised the question that “*How large can the bound in (2.1) be if  $p(z)$  has  $k$  zeros on or outside the unit circle?*” In this connection, Giroux and Rahman [33] proved

**THEOREM 2.2** *For every positive integer  $n$ , there exists a polynomial  $p(z)$  of degree  $n$  having a zero on  $|z| = 1$ , such that*

$$\max_{|z|=1} |p'(z)| \geq (n - c/n) \max_{|z|=1} |p(z)|. \quad (2.3)$$

On the other hand for an arbitrary polynomial  $p(z)$  of degree  $n$  having a zero on  $|z| = 1$ , they proved

$$\max_{|z|=1} |p'(z)| \leq \left(n - \frac{1 - \sin 1}{4\pi n}\right) \max_{|z|=1} |p(z)|. \quad (2.4)$$

Also, it was proposed by R. P. Boas, Jr. to obtain inequalities analogous to (2.1) for polynomials having no zeros in  $|z| < K$ ,  $K > 0$  and the following partial result in this connection was proved by Malik [36].

**THEOREM 2.3** *If  $p(z)$  is a polynomial of degree at most  $n$  having no zeros in  $|z| < K$ ,  $K \geq 1$ , then*

$$\max_{|z|=1} |p'(z)| \leq \left(\frac{n}{1 + K}\right) \max_{|z|=1} |p(z)|. \quad (2.5)$$

*The result is best possible with equality holding for  $p(z) = (z + K)^n$ .*

Govil and Rahman [31] generalized Theorem 2.3 of Malik [36] for any order derivative of the polynomial  $p(z)$  and proved

**THEOREM 2.4** *If  $p(z)$  is a polynomial of degree at most  $n$ , having no zeros in  $|z| < K$ ,  $K \geq 1$ , then*

$$\max_{|z|=1} |p^{(s)}(z)| \leq \frac{n(n-1) \cdots (n-s+1)}{1 + K^s} \max_{|z|=1} |p(z)|. \quad (2.6)$$

For  $s = 1$ , (2.6) obviously reduces to (2.5).

Another generalization of (2.2) was later given by Chan and Malik [10], who proved

**THEOREM 2.5** *If  $p(z) = a_0 + \sum_{\nu=\mu}^n a_\nu z^\nu$  is a polynomial of degree at most  $n$ ,  $p(z) \neq 0$  in  $|z| < K$ ,  $K \geq 1$ , then*

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{1 + K^\mu} \max_{|z|=1} |p(z)|. \quad (2.7)$$



## Self-Reciprocal Polynomials

The equality in (2.7) is attained for  $p(z) = (z^\mu + K^\mu)^{n/\mu}$ ,  $n$  being a multiple of  $\mu$ .

The inequality (2.7) in the case  $\mu = 2$  can also be found in Govil [24].

For some time it was believed that if  $p(z) \neq 0$  in  $|z| < K$ ,  $K \leq 1$ , then the inequality analogous to (2.5) should be

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{1 + K^n} \max_{|z|=1} |p(z)|, \quad (2.8)$$

till E. B. Saff gave the example  $p(z) = (z - \frac{1}{2})(z + \frac{1}{3})$  to counter this belief. As can be easily verified for this polynomial, the left hand side of (2.8) is approximately 2.1666 while the right hand side is  $\frac{2}{(1+(\frac{1}{3})^2)} \max_{|z|=1} |p(z)| \approx 2.144 < 2.166$  and so (2.8) cannot hold true.

Thus the problem in the case  $p(z) \neq 0$  for  $|z| < K$ ,  $K < 1$  is still open. However, for polynomials having all their zeros in  $|z| \leq K$ , the problem is solved for all  $K > 0$ , and following result in this direction is due to Govil [25].

**THEOREM 2.6** *If  $p(z)$  is a polynomial of degree  $n$  having all its zeros in  $|z| \leq K$ , then*

$$\max_{|z|=1} |p'(z)| \geq \begin{cases} \frac{n}{1 + K} \max_{|z|=1} |p(z)| & \text{if } K \leq 1 \\ \frac{n}{1 + K^n} \max_{|z|=1} |p(z)| & \text{if } K \geq 1 \end{cases}$$

*Both the above inequalities are best possible.*

The case  $K < 1$  of the above theorem was also proved by Malik [36], and the case  $K = 1$  by Turán [54]. A simpler proof of the above result in case  $K > 1$  was provided by Datt [11].

If  $p(z)$  is a polynomial of degree  $n$ , it is obviously of interest to obtain an inequality analogous to Bernstein's inequality for polynomials satisfying  $p(z) \equiv z^n p(1/\bar{z})$  or  $p(z) \equiv z^n \overline{p(1/\bar{z})}$ .

The polynomials satisfying  $p(z) \equiv z^n \overline{p(1/\bar{z})}$  are known as self-conjugate polynomials, and for this class of polynomials the following result (see Govil [25], O'Hara and Rodriguez [39] and, Saff and Sheil-Small [50]) is well known.

**THEOREM 2.7** *If  $p(z)$  is a polynomial of degree  $n$  satisfying  $p(z) \equiv z^n \overline{p(1/\bar{z})}$ , then*

$$\max_{|z|=1} |p'(z)| = \frac{n}{2} \max_{|z|=1} |p(z)|. \quad (2.9)$$

Let  $\Pi_n$  denote the class of polynomials of degree  $n$  satisfying  $p(z) \equiv z^n p(1/\bar{z})$ . This class is interesting because for any polynomial  $p(z)$  of degree

N.K. Govil, et al.,

$n$ , the polynomial  $P(z) = z^n p(z + 1/z)$  is always in  $\Pi_{2n}$ . A polynomial belonging to the class  $\Pi_n$  is known as a self-reciprocal polynomial.

It was proposed by Professor Q. I. Rahman to obtain inequality analogous to Bernstein's inequality (2.1) for polynomials belonging to  $\Pi_n$ , and in an attempt to answer this question Govil, Jain and Labelle [29] proved the following partial result.

**THEOREM 2.8** *If  $p(z)$  is a polynomial belonging to  $\Pi_n$  and having all its zeros in the left half-plane or in the right half-plane, then*

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{\sqrt{2}} \max_{|z|=1} |p(z)|. \quad (2.10)$$

*It is not known if (2.10) is best possible, however by considering  $p(z) = z^n + 2iz^{n/2} + 1$ ,  $n$  being even, they showed that if  $p(z)$  simply belongs to  $\Pi_n$ , then the bound in (2.10) can not in general be smaller than  $n/\sqrt{2}$ .*

In fact Ruscheweyh considered the polynomial  $p(z) = (1+iz)^2 + z^{n-2}(z+i)^2$ . Note that this polynomial belongs to  $\Pi_n$  and on  $|z| = 1$ ,  $|p(z)| \leq |1+iz|^2 + |z+i|^2 = |-i+z|^2 + |z+i|^2 = 4$ , while  $|p'(i)| = 4(n-1)$  and so  $\max_{|z|=1} |p'(z)| \geq 4(n-1)$ . Thus  $\max_{|z|=1} |p'(z)| / \max_{|z|=1} |p(z)| \geq (4(n-1))/4 = (n-1) > n/\sqrt{2}$ , implying that if  $p(z)$  only belongs to  $\Pi_n$ , the bound in (2.10) should be something greater than  $n/\sqrt{2}$ .

Aziz [1] considered another subclass of  $\Pi_n$  and proved

**THEOREM 2.9** *Let  $p(z) = \sum_{\nu=0}^n (\alpha_\nu + i\beta_\nu)z^\nu$ ,  $\alpha_\nu \geq 0, \beta_\nu \geq 0$ ,  $\nu = 0, 1, 2, \dots, n$  be a polynomial belonging to  $\Pi_n$ . Then*

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{\sqrt{2}} \max_{|z|=1} |p(z)|. \quad (2.11)$$

*The equality in (2.11) again holds for the polynomial  $p(z) = z^n + 2iz^{n/2} + 1$ ,  $n$  being even.*

As is easy to observe, the hypothesis of Theorem 2.9 is equivalent to that  $p(z)$  belongs to  $\Pi_n$  and that all the coefficients of  $p(z) = \sum_{\nu=0}^n a_\nu z^\nu$  lie in the first quadrant of the complex plane. In fact, if all the coefficients of a polynomial  $p(z)$  belonging to  $\Pi_n$  lie in a sector of opening  $\pi/2$ , say in,  $\psi \leq \arg z \leq \psi + \pi/2$ , for some real  $\psi$ , then the polynomial  $P(z) = e^{-i\psi} p(z)$  belongs to  $\Pi_n$  and has all its coefficients lying in the first quadrant of the complex plane. Since  $\max_{|z|=1} |P(z)| = \max_{|z|=1} |p(z)|$  and  $\max_{|z|=1} |P'(z)| = \max_{|z|=1} |p'(z)|$  we may apply Theorem 2.9 to  $P(z)$  to get that if  $p(z) \in \Pi_n$  and has all its coefficients lying in a sector of opening at most  $\pi/2$ , then also (2.11) holds. The following result that is equivalent to this statement appears in Jain [34].

## Self-Reciprocal Polynomials

**THEOREM 2.10** Let  $p(z) = \sum_{\nu=0}^n a_\nu z^\nu$  where  $a_\nu = \alpha_\nu e^{i\phi} + \beta_\nu e^{i\psi}$ ,  $\alpha_\nu \geq 0$ ,  $\beta_\nu \geq 0$ ,  $1, 2, \dots, n$ ,  $0 \leq |\phi - \psi| \leq \pi/2$ , be a polynomial of degree  $n$ . If further  $p(z) \in \Pi_n$ , then

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{\sqrt{2}} \max_{|z|=1} |p(z)|.$$

The result is best possible with equality holding for the polynomial  $p(z) = z^n + 2iz^{n/2} + 1$ ,  $n$  being an even integer.

Later, Datt and Govil [12] proved the following result which can yield both the above results Theorem 2.9 and Theorem 2.10.

**THEOREM 2.11** Let  $p(z) = \sum_{\nu=0}^n (\alpha_\nu + i\beta_\nu)z^\nu$  be a polynomial of degree  $n$  belonging to  $\Pi_n$ . If on  $|z| = 1$ , the maximum of  $|\sum_{\nu=0}^n \alpha_\nu z^\nu|$  and  $|\sum_{\nu=0}^n \beta_\nu z^\nu|$  is attained at the same point, then

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{\sqrt{2}} \max_{|z|=1} |p(z)|.$$

The equality here holds again for  $p(z) = z^n + 2iz^{n/2} + 1$ ,  $n$  being an even integer.

Govil and Vetterlein [32] obtained a bound for  $\max_{|z|=1} |p'(z)|$  which depends on the opening of the sector containing all the coefficients of a self-reciprocal polynomial and includes as special cases Theorem 2.9 and 2.10. Further, their result is applicable even when the opening of the sector is greater than  $\pi/2$ . More precisely, their result is

**THEOREM 2.12** Let  $p(z) = \sum_{\nu=0}^n a_\nu z^\nu$  is a polynomial belonging to  $\Pi_n$ , with its coefficients lying in a sector of opening  $\gamma$  with vertex at the origin where  $0 \leq \gamma \leq 2\pi/3$ , then

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{2 \cos(\gamma/2)} \max_{|z|=1} |p(z)|.$$

The result is best possible with equality holding for the polynomial  $p(z) = z^n + 2iz^{n/2} + 1$ ,  $n$  being an even integer.

Rahman and Tariq [44] observed that in the above theorem a sharp estimate of  $\max_{|z|=1} |p'(z)|$ , that is valid for  $\gamma$  in  $[0, \pi)$  instead of  $[0, 2\pi/3]$ , can be given in terms of  $|p(1)|$  and that the proof given by Govil and Vetterlein [32] can be easily modified to prove the following

**THEOREM 2.13** Let  $p(z) = \sum_{\nu=0}^n a_\nu z^\nu$  be a polynomial belonging to  $\Pi_n$ , with its coefficients lying in a sector of opening  $\gamma$  with vertex at the origin where  $0 \leq \gamma < \pi$ , then

$$\max_{|z|=1} |p'(z)| \leq \frac{n}{2 \cos(\gamma/2)} |p(1)|. \quad (2.12)$$

N.K. Govil, et al.,

*In the case where  $n$  is even, the polynomial  $p(z) := z^n + 2e^{i\gamma}z^{n/2} + 1$  shows that the above inequality is sharp for any  $\gamma \in [0, \pi)$ .*

Although the class  $\Pi_n$  of polynomials has been extensively studied among others by Frappier and Rahman [20] and Frappier, Rahman and Ruscheweyh [21], the problem of obtaining a sharp inequality analogous to Bernstein's inequality (2.1) is still open for polynomials of degree  $n \geq 3$ . However, the following sharp inequality in the reverse direction, which is easy to obtain, is due to Dewan and Govil [14].

**THEOREM 2.14** *If  $p(z)$  is a polynomial belonging to  $\Pi_n$ , then*

$$\max_{|z|=1} |p'(z)| \geq \frac{n}{2} \max_{|z|=1} |p(z)|. \quad (2.13)$$

*The result is best possible and the equality holds for  $p(z) = (z^n + 1)$ .*

It may be remarked that several of the above mentioned results for polynomials have been extended to entire functions of exponential type, and this, we take up in the next section.

### 3 Entire functions of exponential type and uniformly almost periodic functions

In this section we will discuss the extension of some of the results on polynomials mentioned in Section 2 to entire functions of exponential type, and of Theorem 2.13 to entire functions of exponential type that are uniformly almost periodic on the real line. We start with the following definitions.

Let  $f$  be an entire function and  $r$  be any positive real number. We will denote

$$M(r) = M_f(r) := \max_{|z|=r} |f(z)|.$$

The order (or the order of growth) of an entire function  $f$ , denoted by  $\rho$ , is defined by

$$\rho := \limsup_{r \rightarrow \infty} \frac{\log \log M(r)}{\log r}.$$

It is a convention to take the order of a constant function of modulus less than or equal to one as 0. An entire function of finite order  $\rho$  is said to have type  $T$ , where  $T$  is given by

$$T := \limsup_{r \rightarrow \infty} \frac{\log M(r)}{r^\rho}.$$

## Self-Reciprocal Polynomials

**DEFINITION 1** An entire function  $f$  is said to be of exponential type  $\tau$  if for every  $\varepsilon > 0$  there is a constant  $k(\varepsilon)$  such that  $|f(z)| \leq k(\varepsilon) e^{(\tau+\varepsilon)|z|}$  for all  $z \in \mathbb{C}$ .

It is clear that entire functions of order less than 1 are of exponential type  $\tau$ , where  $\tau$  can be taken to be any number greater than or equal to 0. Also entire functions of order 1 and type  $T \leq \tau$  are of exponential type  $\tau$ .

Examples of entire functions of exponential type includes polynomials with complex coefficients,  $\sin \tau z$ ,  $\cos \tau z$  etc.

**DEFINITION 2** Let  $f$  be an entire function of exponential type. The function

$$h_f(\theta) := \limsup_{r \rightarrow \infty} \frac{\log |f(re^{i\theta})|}{r}, \quad 0 \leq \theta < 2\pi. \quad (3.1)$$

is called the indicator function of  $f$ . It describes the growth of the function along a ray  $\{z | \arg z = \theta\}$ . It is finite or  $-\infty$ . Unless  $h_f(\theta) \equiv -\infty$ , it is a continuous function of  $\theta$ .

Bernstein himself (see [3], p. 102) found the extension of inequality (2.1) for the entire functions of exponential type that are bounded on the real line. He in fact proved

**THEOREM 3.1** *Let  $f$  be an entire function of exponential type  $\tau$ , bounded on the real axis. Then*

$$\sup_{-\infty < x < \infty} |f'(x)| \leq \tau \sup_{-\infty < x < \infty} |f(x)|. \quad (3.2)$$

*The estimate is sharp.*

Equality holds for  $f(z) = \sin \tau z$ . In fact equality holds if and only if  $f(z) = a e^{i\tau z} + b e^{-i\tau z}$ , where  $a, b \in \mathbb{C}$  and  $|a| + |b| > 0$ .

If  $p(z) := \sum_{\nu=0}^n c_\nu z^\nu$  is a polynomial of degree at most  $n$  with complex coefficients, then  $f(z) = p(e^{iz})$  is an entire function of exponential type  $n$ . Furthermore, if we assume that  $p(z)$  has no zero in the open unit disk  $|z| < 1$  then  $f(z)$  will have no zero in the open half plane  $\text{Im}(z) > 0$ . Also

$$h_f\left(\frac{\pi}{2}\right) := \limsup_{y \rightarrow \infty} \frac{\log |f(iy)|}{y} = 0$$

In the light of above and Theorem 2.1, the following result of Boas [8], see also [40], provides the generalization of Theorem 2.1 for entire functions of exponential type.

**THEOREM 3.2** *Let  $f$  be an entire function of exponential type  $\tau$ , bounded on the real axis,  $f(z) \neq 0$  for  $y > 0$ , and  $h_f(\pi/2) = 0$ . Then*

$$\sup_{-\infty < x < \infty} |f'(x)| \leq \frac{\tau}{2} \sup_{-\infty < x < \infty} |f(x)|. \quad (3.3)$$

N.K. Govil, et al.,

*The inequality is sharp.*

The function  $f(x) = (1 + e^{i\tau z})/2$  serves as an extremal function for the inequality. We observe that  $|f(x)| = |(1 + e^{i\tau z})/2| \leq 1$  for all  $x \in \mathbb{R}$  and  $f(0) = 1$ . So  $\sup_{-\infty < x < \infty} |f(x)| = 1$ .  $f(x) = 0$  if and only if  $x = (2n + 1)\pi/\tau$ , where  $n$  is any integer. Thus  $f$  has only real zeros.

$$h_f\left(\frac{\pi}{2}\right) = \limsup_{y \rightarrow \infty} \frac{\log |f(iy)|}{y} = \limsup_{y \rightarrow \infty} \frac{\log\left(\frac{1+e^{-\tau y}}{2}\right)}{y} = 0$$

and

$$|f'(x)| = \left| \frac{i\tau e^{i\tau x}}{2} \right| = \frac{\tau}{2} \quad \text{for all } x.$$

Thus

$$\sup_{-\infty < x < \infty} |f'(x)| = \frac{\tau}{2} \sup_{-\infty < x < \infty} |f(x)|.$$

It is worth mentioning that the condition  $h_f(\frac{\pi}{2}) = 0$  in the above theorem is necessary as it can be seen from the function  $f(x) = \cos \tau z$  which is an entire function of exponential type  $\tau$  having all the zeros on the real axis. That is  $f(z) \neq 0$  for  $y > 0$ ,  $\sup_{-\infty < x < \infty} |f(x)| = 1$ ,

$$h_f\left(\frac{\pi}{2}\right) = \limsup_{y \rightarrow \infty} \frac{\log\left(\frac{e^{-\tau y} + e^{\tau y}}{2}\right)}{y} = \tau > 0$$

and

$$|f'(x)| = \tau = \tau \sup_{-\infty < x < \infty} |f(x)| \quad \text{for all } x \in \mathbb{R}.$$

It was proved by Rahman [41] that the equality holds in (3.3) for any entire function of exponential type  $\tau > 0$  satisfying conditions of the Theorem 3.2 whose zeros are all real.

Govil and Jain [28] studied class of entire functions of exponential type  $\tau > 0$  satisfying  $f(z) \equiv e^{i\tau z} \overline{f(\bar{z})}$ . They proved the following theorem for this class.

**THEOREM 3.3** *If  $f$  is an entire function of exponential type  $\tau$  such that  $f(z) \equiv e^{i\tau z} \overline{f(\bar{z})}$ , then*

$$\sup_{-\infty < x < \infty} |f'(x)| = \frac{\tau}{2} \sup_{-\infty < x < \infty} |f(x)|. \quad (3.4)$$

The above result provides an extension of Theorem 2.7 for entire functions of exponential type.

## Self-Reciprocal Polynomials

It is clear that if  $z_0$  is a zero of  $f$  that satisfies the condition  $f(z) \equiv e^{i\tau z} \overline{f(\bar{z})}$ , then so is  $\bar{z}_0$ . Thus upper half plane  $Im(z) > 0$  has as many zeros of  $f$  as the lower half plane  $Im(z) < 0$  has.

To see, what to expect as an extension of self-reciprocal polynomials, consider  $f(z) = p(e^{iz})$ , where  $p(z)$  is a self-reciprocal polynomial of degree  $n$ . The function  $f$  is an entire function of exponential type  $n$  satisfying the condition

$$f(z) \equiv e^{inz} f(-z).$$

This led Govil [27] to consider a class of entire functions of exponential type  $\tau$  that satisfies the condition  $f(z) \equiv e^{i\tau z} f(-z)$ . He proved several results for functions in this class. One of his result which provides an extension of Theorem 2.14 for entire functions of exponential type is

**THEOREM 3.4** *Let  $f$  be an entire function of exponential type  $\tau$ , satisfying the condition  $f(z) \equiv e^{i\tau z} f(-z)$ . Then*

$$\sup_{-\infty < x < \infty} |f'(x)| \geq \frac{\tau}{2} \sup_{-\infty < x < \infty} |f(x)|. \quad (3.5)$$

*The result is best possible and the equality holds for  $f(z) = (1 + e^{i\tau z})$ .*

He deduced Theorem 3.4 as a corollary of another inequality that he proved for the functions in this class. However, the result can also be proved directly, and the proof is not difficult.

Let us denote the ratio  $\sup_{-\infty < x < \infty} |f'(x)| / \sup_{-\infty < x < \infty} |f(x)|$  by  $q(f)$ . Theorem 3.4 gives us the smallest value of  $q(f)$  if  $f$  satisfies the condition  $f(z) \equiv e^{i\tau z} f(-z)$ . From Theorem 3.1, we already know that,  $q(f)$  could be as large as  $\tau$  and it will be equal to  $\tau$  only when the function is of the form  $ae^{i\tau z} + be^{-i\tau z}$ . However, it is easy to see that functions satisfying the condition  $f(z) \equiv e^{i\tau z} f(-z)$  can not be of this form. So, again from Theorem 3.1,  $q(f) < \tau$  for such functions. Thus the question is: "How large  $q(f)$  could be for such functions"?

Rahman and Tariq [45] have shown through an example that it could be as close to  $\tau$  as one wish. In fact the following result holds true.

**THEOREM 3.5** *Given any number  $\varepsilon \in (0, \tau)$ , we can find an entire function  $f_\varepsilon$  of exponential type  $\tau$  which satisfies the condition  $f_\varepsilon(z) \equiv e^{i\tau z} f_\varepsilon(-z)$  such that*

$$\sup_{-\infty < x < \infty} |f'_\varepsilon(x)| \geq (\tau - \varepsilon) \sup_{-\infty < x < \infty} |f_\varepsilon(x)|.$$

It is clear that if  $z_0$  is a zero of  $f$  that satisfies the condition  $f(z) \equiv e^{i\tau z} f(-z)$ , then so is  $-z_0$ . Thus upper half plane  $Im(z) > 0$  has as many zeros of  $f$  as the lower half plane  $Im(z) < 0$  has.

N.K. Govil, et al.,

In view of the above observation, this result is a bit surprising because functions satisfying condition  $f(z) \equiv e^{i\tau z} f(-z)$  are in some sense similar to those discussed in Theorem 3.3. They both have as many zeros in the upper half plane as in the lower half plane.

Rahman and Tariq [45] formulated and proved an extension of the result of Govil and Vetterlien [32] to the entire function of exponential type. Theorem 2.13 requires the coefficients of a self-reciprocal polynomial to lie in certain sector. The main issue while deciding about the extension to the entire function of exponential type was: "What class of entire functions of exponential type would admit an extension of Theorem 2.13"?

If we simply take functions of the form  $f(z) = p(e^{iz}) = \sum_{\nu=0}^n a_{\nu} e^{i\nu z}$  and require coefficients to lie in a sector, then it is indeed an entire function of exponential type but too restrictive as an arbitrary entire function of exponential type, in general, cannot be expressed as a finite or infinite sum of the form  $\sum a_{\nu} e^{i\nu z}$ . According to Rahman and Tariq [45] an appropriate class of entire functions of exponential type for which the Theorem 2.13 will admit an extension is the one whose elements are uniformly almost periodic on the real line. Under certain conditions, functions that are uniformly almost periodic on the real line may be represented as a sum  $\sum a_{\nu} e^{i\lambda_{\nu} z}$ . They gave the following extension of Theorem 2.13 for entire functions of exponential type.

**THEOREM 3.6** *Let  $f$  be an entire function of exponential type  $\tau$  such that  $f(z) \equiv e^{i\tau z} f(-z)$ . Furthermore, let  $f$  be uniformly almost periodic on the real axis, with Fourier series  $f(x) \sim \sum_{n=1}^{\infty} A_n e^{i\Lambda_n x}$ , where the coefficients  $A_1, A_2, \dots$  lie in a sector of opening  $\gamma \in [0, \pi)$  with vertex at the origin. Then*

$$\sup_{-\infty < x < \infty} |f'(x)| \leq \frac{\tau}{2 \cos(\gamma/2)} |f(0)|. \quad (3.6)$$

*The example  $f(z) := e^{i\tau z} + 2e^{i\gamma} e^{i\tau z/2} + 1$  shows that the estimate is sharp.*

In the rest of the paper, we will present a brief overview of the relevant part of the theory of uniformly almost periodic functions that is needed for the understanding of Theorem 3.6. For the detail studies, we refer readers to [5], [9] and [45].

Let  $f : \mathbb{R} \rightarrow \mathbb{C}$  be a continuous function, and  $\varepsilon$  an arbitrary positive number. A real number  $t = t(\varepsilon) = t_f(\varepsilon)$  is called a *translation number* of  $f$  corresponding to  $\varepsilon$  provided that  $|f(x+t) - f(x)| \leq \varepsilon$  for all real  $x$ .

A set  $E$  of real numbers  $t$  is called *relatively dense* if there exists a positive number  $L$  such that any interval  $(\alpha, \alpha + L)$  of length  $L$  contains at least one number  $t$  of the set  $E$ .



## Self-Reciprocal Polynomials

**DEFINITION 3** A continuous function  $f : \mathbb{R} \rightarrow \mathbb{C}$  is called *uniformly almost periodic*, *u. a. p.* for short, if there exists a relatively dense set of translation numbers of  $f$  corresponding to any given  $\varepsilon > 0$ . In other words, for any  $\varepsilon > 0$  we can find a positive number  $L = L(\varepsilon)$  such that an arbitrary interval of length  $L$  contains at least one translation number  $t(\varepsilon)$ .

It is easy to see that if  $f$  is *u.a.p.* then so is  $cf$  and  $f(x+c)$  for any constant  $c$ ; since  $||f(x+t)| - |f(x)|| \leq |f(x+t) - f(x)|$ ,  $|f|$  is *u.a.p.* if  $f$  is; the sum and the product of *u.a.p.* functions are *u.a.p.* [5, pp. 2–6, 12–15].

(i) Every periodic function is *u. a. p.* (ii) *u.a.p.* functions are bounded on the real line. (iii) A *u.a.p.* function  $f$  is uniformly continuous for  $-\infty < x < \infty$ . (iv) The limit function of a uniformly convergent sequence of *u.a.p.* functions on  $(-\infty, \infty)$  is also *u.a.p.* (v) For any *u.a.p.* function  $f$ ,  $\lim_{T \rightarrow \infty} (1/T) \int_0^T f(x) dx$  exists and is finite. It is called the mean value of the function  $f$  and is denoted by  $\mathcal{M}\{f\}$ . Since  $|f|$  is *u.a.p.* if  $f$  is, it implies that  $\mathcal{M}\{|f|\}$  also exists for a *u.a.p.* function  $f$ .

For any  $\lambda \in \mathbb{R}$  and  $f$  a *u.a.p.* function,  $f(x)e^{-i\lambda x}$  is a *u.a.p.* function. Hence

$$a(\lambda) = \mathcal{M}\{f(x)e^{-i\lambda x}\} := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x)e^{-i\lambda x} dx \quad (3.7)$$

exists.

The following result about  $a(\lambda)$  [5, pp. 16–18] plays an important role in the development of the theory of Fourier series of *u.a.p.* functions.

**THEOREM 3.7** Let  $f$  be a *u. a. p.* function, and  $a(\lambda)$  as in (3.7). Then  $a(\lambda) = 0$  except for a countable number of  $\lambda$ 's.

For entire functions of exponential type bounded on real line, Rahman & Tariq [45] have proved the following

**THEOREM 3.8** If  $f$  is an entire function of exponential type  $\tau$  bounded on the real axis, and let  $\lambda$  is any real number such that  $|\lambda| > \tau$ . Then

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(x)e^{i\lambda x} dx = 0.$$

So, from the above two theorems we conclude that, if  $f$  is an entire function of exponential type  $\tau$ , and *u.a.p.* on the real line, then  $a(\lambda) = 0$  except for countably many  $\lambda$ 's such that  $|\lambda| \leq \tau$ .

For a *u.a.p.*  $f$ , the values of  $\lambda$  for which  $a(\lambda) \neq 0$  are called *Fourier exponents* and the corresponding  $a(\lambda)$ 's *Fourier coefficients* of the function.

N.K. Govil, et al.,

If  $\Lambda_1, \Lambda_2, \dots$ , are the Fourier exponents and  $A_n := a(\Lambda_n)$  for  $n = 1, 2, 3, \dots$  corresponding Fourier coefficients, the formal series

$$\sum_{n=1}^{\infty} A_n e^{i\Lambda_n x} = A_1 e^{i\Lambda_1 x} + A_2 e^{i\Lambda_2 x} + \dots$$

is called the Fourier series of  $f$ , and we write  $f(x) \sim \sum_{n=1}^{\infty} A_n e^{i\Lambda_n x}$ .

An important theorem about the Fourier series of a *u.a.p.* function is the following *uniqueness theorem*.

**THEOREM 3.9** *A u. a. p. is always uniquely determined by its Fourier series.*

That is, if  $f_1$  and  $f_2$  are two distinct *u.a.p.* functions then their Fourier series will be different. The next result [5, pp. 21–28] is called "Bohr's fundamental theorem" or "Parseval equation" which is equivalent to uniqueness theorem [9].

**THEOREM 3.10** *Let  $f(x) \sim \sum_{n=1}^{\infty} A_n e^{i\Lambda_n x}$  be any u.a.p. function with its Fourier series. Then*

$$\mathcal{M}\{|f|^2\} = \sum_{n=1}^{\infty} |A_n|^2. \quad (3.8)$$

It is easy to see that every trigonometric polynomial is *u.a.p.*. Since the limit of a uniformly convergent sequence of *u.a.p.* functions is *u.a.p.*, the limit of a uniformly convergent sequence of trigonometric polynomials is also a *u.a.p.* function. The converse of this result is also true and is contained in the following Approximation Theorem.

**THEOREM 3.11** *Given a u.a.p. function  $f(x) \sim \sum_{n=1}^{\infty} A_n e^{i\Lambda_n x}$  and a positive number  $\varepsilon$ , there exist a trigonometric polynomial  $P(x)$  whose exponents are Fourier exponents of  $f$  and which satisfies  $|f(x) - P(x)| < \varepsilon$  for all  $x$  in  $\mathbb{R}$ .*

H. Bohr [9] gave the proof of this important theorem. Later on, S. Bochner also proved it by giving an elegant extension of Fejer sum to the class of *u.a.p.* functions. We refer readers to [5], [9], [45] for the proof and other details.

In general, the trigonometric polynomials determined by the partial sums of the Fourier series of a function may not converge uniformly to the function for all  $x$  in  $\mathbb{R}$ . That is, Fourier series of a function may not represent the function in general. It is useful to have conditions under which the Fourier series does represent the function. One such condition is given in the next

## Self-Reciprocal Polynomials

theorem. But before we discuss it, we need to define what is meant by linearly independent set of numbers.

**DEFINITION 4** A set  $\alpha_\nu$  ( $\nu = 1, 2, \dots$ ) of real numbers is called *linearly independent* if for any  $n$  the only rational non zero values of  $r_1, r_2, \dots, r_n$  satisfying the equation  $r_1\alpha_1 + r_2\alpha_2 + \dots + r_n\alpha_n = 0$  are  $r_1 = r_2 = \dots = r_n = 0$ .

**THEOREM 3.12** For a uniformly almost periodic function  $f$  with linearly independent Fourier exponents, the Fourier series  $\sum_{n=1}^{\infty} A_n e^{i\Lambda_n x}$  converges uniformly for all  $x$ .

In fact, the series formed by the absolute values  $\sum_{n=1}^{\infty} |A_n|$  converges. By virtue of the uniqueness theorem  $f(x) = \sum_{n=1}^{\infty} A_n e^{i\Lambda_n x}$ . It is also known [5, pp. 51–52] that the Fourier series  $\sum_{n=1}^{\infty} A_n e^{i\Lambda_n x}$  of a *u.a.p.* function is absolutely convergent if the Fourier coefficients  $A_1, A_2, \dots$  are positive.

We conclude the paper by presenting the following result due to Rahman and Tariq [45], which gives a condition on the Fourier coefficients so that the Fourier series represent the function. In fact this is the main result of their paper [45] which has been used to prove Theorem 3.6 that provides the extension of the theorem of Govil and Vetterlien [32] about self-reciprocal polynomials to the entire functions of exponential type.

**THEOREM 3.13** Let  $f$  be a *u.a.p.* function

$$f(x) \sim \sum_{n=1}^{\infty} A_n e^{i\Lambda_n x},$$

where the Fourier coefficients  $A_1, A_2, \dots$  lie in a sector of opening  $\gamma \in [0, \pi)$  with vertex at the origin. Then  $\sum_{n=1}^{\infty} |A_n| < \infty$ .

## References

- [1] A. Aziz, Inequalities for the derivative of a polynomial, *Proc. Amer. Math. Soc.* **89** (1983), 259-266.
- [2] A. Aziz and Q. G. Mohammad, A simple proof of a theorem of Erdős and Lax, *Proc. Amer. Math. Soc.* **80** (1980), 119-122.
- [3] S.N. Bernstein, *Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle*, Gauthier-Villars, Paris, 1926.

N.K. Govil, et al.,

- [4] S. Bernstein, Sur L'ordre de la meilleur approximation des fonctions continues par des polynômes de degré donné, *Memoire de l'Académie Royal de Belgique* (2) **4** (1912), 1 - 103.
- [5] A.S. Besicovitch, Almost Periodic Functions, Dover Publications, Inc., 1954.
- [6] R.P. Boas, Jr., Entire functions, Academic Press, New York, 1954.
- [7] R. P. Boas, Inequalities for the derivatives of polynomials, *Mathematics Magazine* **42** (1969), 165-174.
- [8] R. P. Boas, Inequalities for asymmetric entire functions, *Illinois J. Math.* **1** (1957), 94-97.
- [9] H. Bohr, Almost Periodic Functions, Chelsea Publishing Company, New York, 1947.
- [10] T. N. Chan and M. A. Malik, On Erdős-Lax Theorem, *Proc. Indian Acad. Sci.* **92** (1983), 191-193.
- [11] B. Datt, A Note on the derivative of a polynomial, *Math. Student* **43** (1975), 299-300.
- [12] B. Datt and N.K. Govil, Some inequalities for polynomials satisfying  $p(z) \equiv z^n p(1/z)$ , *Approx. Theory & Appl. (N. S.)* **12** (1996), 40-44.
- [13] N. G. de Bruijn, Inequalities concerning polynomials in the complex domain, *Neder. Akad. Wetensch. Proc.* **50** (1947), 1265-1272.
- [14] K.K. Dewan and N.K. Govil, An inequality for the derivative of self-inversive polynomials, *Bull. Austral. Math. Soc.*, **27** (1983), 403-406.
- [15] A. Durand, *Quelques aspects de la theorie analytique des polynômes, I et II*, Université de Limoges, 1984.
- [16] P. Erdős, On extremal properties of the derivatives of polynomials, *Ann. of Math* (2) **41** (1940), 310-313.
- [17] L. Fejer, Über konjugierte trigonometrische Reihen, *J. Reine Angew Math.* **144** (1914), 48-56.
- [18] L. Fejer, Über einen S. Bernsteinschen Satz über die Derivwerte eines trigonometrischen Polynoms und über die Szegösche Verschärfung desselben, *Bull. Calcutta Math. Soc.* **20** (1930), 49-54.

- [19] M. Feketé, Über einen Satz des Herrn Serge Bernstein, *J. Reine Angew. Math.* **146** (1916), 88-94.
- [20] C. Frappier and Q. I. Rahman, On an inequality of S. Bernstein, *Canad. J. Math.* **34** (1982), 932-944.
- [21] C. Frappier, Q. I. Rahman and St. Ruscheweyh, New inequalities for polynomials, *Trans. Amer. Math. Soc.* **288** (1985), 69-99.
- [22] C. Frappier, Q.I. Rahman and St. Ruscheweyh, Inequalities for polynomials, *J. Approx. Theory*, **44** (1985), 73-81.
- [23] R. Gardner and N. K. Govil, Some inequalities for entire functions of exponential type, *Proc. Amer. Math. Soc.* **123** (1995), 2757-2761.
- [24] N. K. Govil, On the maximum modulus of polynomials, *J. Math. Anal. Appl.* **112** (1985), 253-258.
- [25] N. K. Govil, On the derivative of a polynomial, *Proc. Amer. Math. Soc.* **41** (1973), 543-546.
- [26] N.K. Govil, On maximum modulus of polynomials satisfying  $p(z) \equiv z^n p(1/z)$ , *East J. Approx.*, **3** (1997), 111-115.
- [27] N.K. Govil,  $L^p$  Inequalities for entire functions of exponential type, *Math. Ineq. & Appl.*, **6** (2003), 445-452.
- [28] N.K. Govil and V.K. Jain, An integral inequality for functions of exponential type, *Annals. Univ. Marie Curie Skłodowska* **39** (1985), 57-60.
- [29] N. K. Govil, V. K. Jain and G. Labelle, Inequalities for polynomials satisfying  $p(z) \equiv z^n p(\frac{1}{z})$ , *Proc. Amer. Math. Soc.* **57** (1976), 238-242.
- [30] N.K. Govil and R. N. Mohapatra, Markov and Bernstein Type Inequalities for Polynomials, *J. of Inequal. & Appl.* **3** (1999), 349-387.
- [31] N. K. Govil and Q. I. Rahman, Functions of exponential type not vanishing in a half-plane and related polynomials, *Trans. Amer. Math. Soc.* **137** (1969), 501-517.
- [32] N.K. Govil and D.H. Vetterlein, Inequalities for a class of polynomials satisfying  $p(z) \equiv z^n p(1/z)$ , *Complex Variables Theory Appl.* **31** (1996), 185-191.

N.K. Govil, et al.,

- [33] A. Giroux and Q.I. Rahman, Inequalities for Polynomials with a prescribed zero, *Trans. Amer. Math. Soc.* **193** (1974), 67-98.
- [34] V.K. Jain, Inequalities for polynomials satisfying  $p(z) \equiv z^n p(1/z)$  II, *J. Indian Math. Soc.* **59** (1993), 167–170.
- [35] P. D. Lax, Proof of a conjecture of P. Erdős. On the derivative of a polynomial, *Bull. Amer. Math. Soc.* **50** (1944), 509-513.
- [36] M. A. Malik, On the derivative of a polynomial, *J. London Math. Soc.* **1** (1969), 57-60.
- [37] A. A. Markov, On a problem of D. I. Mendelev (Russian), *Zapiski Imp. Akad. Nauk* **62** (1889), 1-24.
- [38] D. Mendelev, Investigation of aqueous solutions based on specific gravity (Russian), St. Petersburg, 1887.
- [39] P. J. O'Hara and R. S. Rodriguez, Some properties of self-inversive polynomials, *Proc. Amer. Math. Soc.* **44** (1974), 331-335.
- [40] Q. I. Rahman, Functions of exponential type, *Trans. Amer. Math. Soc.* **135** (1969), 295-309.
- [41] Q. I. Rahman, On asymmetric entire functions, *Proc. Amer. Math. Soc.* **14** (1963) 507–508.
- [42] Q. I. Rahman and G. Schmeisser, *Les inégalités de Markov et de Bernstein*, Le Presses de l'Université de Montréal, Montréal, Canada, 1983.
- [43] Q.I. Rahman and G. Schmeisser, *Analytic theory of polynomials*, Clarendon Press, Oxford, 2002.
- [44] Q.I. Rahman and Q.M. Tariq, An inequality for 'self-reciprocal' polynomials, *East J. Approx.* **12** (2006) 43–51.
- [45] Q.I. Rahman and Q.M. Tariq, On Bernstein's inequality for entire functions of exponential type, *Comput. Methods Funct. Theory* **7** (2007) 167–184.
- [46] Th. M. Rassias, On certain properties of polynomials and their derivatives, In: *Topics in Mathematical Analysis*, Th. M. Rassias, ed., World Scientific Publishing Company, Singapore, 1989, pp. 758-802.

## Self-Reciprocal Polynomials

- [47] F. Riesz, Sur les polynômes trigonometriques, *Comptes Rendus Acad. Sci. Paris* **158** (1914), 1657-1661.
- [48] M. Riesz, Eine trigonometrische interpolation formel und einige Ungleichung für Polynome, *Jahresbericht der Deutschen Mathematiker-Vereinigung* **23** (1914), 354-368.
- [49] W. W. Rogosinski, Extremal problems for polynomials and trigonometric polynomials, *J. London Math. Soc.* **29** (1954), 259-275.
- [50] E. B. Saff and T. Sheil-Small, Coefficient and integral mean estimates for algebraic and trigonometric polynomials with restricted zeros, *J. London Math. Soc.* **9** (1974), 16-22.
- [51] A. C. Schaeffer, Inequalities of A. Markov and S. Bernstein for polynomials and related functions, *Bull. Amer. Math. Soc.* **47** (1941), 565-579.
- [52] G. Szegő, Über einen Satz des Herrn Serge Bernstein, *Schriften der Königsberger Gelehrten Gesellschaft* **5** (1928), 59-70.
- [53] S. A. Telyakovskii, Research in the theory of approximation of functions at the mathematical institute of the academy of sciences, *Trudi Mat. Inst. Steklov* **182** (1988); English Trans. in *Proc. Steklov Inst. Math.*, 1990 No. **1**, 141-197.
- [54] P. Turán, Über die Ableitung von Polynomen, *Compositio Math.* **7** (1939), 85-95.
- [55] C. de la Vallée Poussin, Sur le maximum du module de la dérivée d'une expression trigonometrique d'ordre et de module bornés, *Comptes Rendus de l'Académie des Sciences, Paris* **166** (1918), 843-846.
- [56] E. V. Voronovskaja, The Functional Method and its Applications, *Trans. Math. Monographs*, Vol. **28**, Amer. Math. Soc., Providence 1970.

## Numerical approximation to $\pi$ using parabolic segments

Mark Bollman  
Department of Mathematics  
and Computer Science  
Albion College  
Albion, MI 49224

George Grossman  
Department of Mathematics  
Central Michigan University  
Mount Pleasant, MI 48858

**Abstract.** We derive numerical algorithms that can be used to approximate  $\pi$ . We utilize and extend standard recurrence relations that compute the area of inner and outer regular polygons. The relations that we derive arise from approximations of area of circular sectors by adjoining parabolic segments to triangular subregions of the regular polygons. We also discuss the accuracy of the new approach and employ Mathematica software in the present work to facilitate computation.

### Preliminary Results

Archimedes employed perimeters of regular polygons to approximate  $\pi$ ; [1] contains a summary of his approach. In the present paper we use Archimedes' results on the area of parabolic segments to extend several results in [2] which utilized areas of regular polygons to approximate  $\pi$ . Let  $p_n$ , respectively  $P_n$ , (see fig. 1,  $n = 3$ ), denote regular polygons inscribing and circumscribing the unit circle, each having  $2^{n+1}$  sides. We note from [2], that both  $P_n$  and  $p_n$  can be subdivided into  $2^{n+1}$  congruent, isosceles triangles each having interior angle  $\pi/2^n$ , at the origin (fig. 4 shows half of triangle). The other two equal angles at the circumference are equal, in radians, to

$$\frac{\pi}{2} \left( 1 - \frac{1}{2^n} \right).$$

It is also noted that the interior angle is halved for successive values of  $n$ . It is known, [2], that from the Taylor series for  $\tan$ ,  $\sin$ , and  $\cos$ , we have

$$\begin{aligned} (1) \quad \text{area}(P_n) &= 2^{n+1} \tan \left( \frac{\pi}{2^{n+1}} \right) \\ &= \pi \left( 1 + \left( \frac{\pi}{2^{n+1}} \right)^2 \frac{1}{3} + \left( \frac{\pi}{2^{n+1}} \right)^4 \frac{2}{15} + \left( \frac{\pi}{2^{n+1}} \right)^6 \frac{17}{315} + \dots \right), \end{aligned}$$

$$\begin{aligned} (2) \quad \text{area}(p_n) &= 2^{n+1} \sin \left( \frac{\pi}{2^{n+1}} \right) \cos \left( \frac{\pi}{2^{n+1}} \right) = 2^n \sin \left( \frac{\pi}{2^n} \right) \\ &= \pi \left( 1 - \left( \frac{\pi}{2^n} \right)^2 \frac{1}{6} + \left( \frac{\pi}{2^n} \right)^4 \frac{1}{120} - \left( \frac{\pi}{2^n} \right)^6 \frac{1}{5040} + \dots \right). \end{aligned}$$

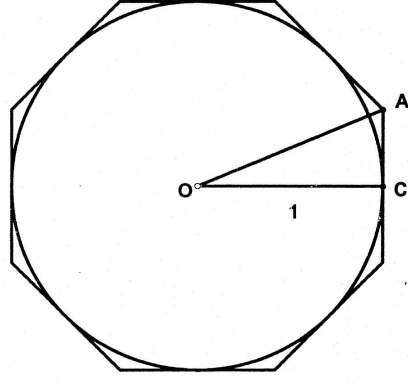
We also have from [2] that

$$(3) \quad \left| \frac{\text{area}(p_n) + 2 \cdot \text{area}(P_n)}{3} - \pi \right| \approx 10^{-1.2n},$$

so that we can write

$$(4) \quad \pi = \frac{1}{3} \cdot \text{area}(p_n) + \frac{2}{3} \cdot \text{area}(P_n) + O(16^{-n}),$$



Figure 1:  $P_2$ , regular polygon of 8 sides.

which follows from the elimination of second term in the power series in (1,2). Similarly, by this process of elimination, we have from (1,2,4) that

$$(5) \quad (\pi) = \frac{4}{3} \cdot \text{area}(p_{n+1}) - \frac{1}{3} \cdot \text{area}(p_n) + O(16^{-n}),$$

$$(6) \quad = \frac{4}{3} \cdot \text{area}(P_{n+1}) - \frac{1}{3} \cdot \text{area}(P_n) + O(16^{-n}),$$

$$(7) \quad = \frac{1}{3} \cdot \text{area}(P_n) + \frac{2}{3} \cdot \text{area}(p_{n+1}) + O(16^{-n}),$$

$$(8) \quad = \frac{1}{3} \cdot \text{area}(p_{n+1}) + \frac{2}{3} \cdot \text{area}(P_{n+1}) + O(16^{-n}),$$

$$(9) \quad = \frac{1}{9} \cdot \text{area}(p_n) + \frac{8}{9} \cdot \text{area}(P_{n+1}) + O(16^{-n}).$$

The foregoing, (5-9) represents the six possible combinations of areas involving pairs of distinct terms from  $\{P_n, p_n, P_{n+1}, p_{n+1}\}$ . We now show how to derive (5,7) using areas of parabolic segments and employing Archimedes' formula. Consider  $\triangle OPR, \triangle PCR$ , (fig. 3.) By Archimedes' well-known formula and approximation we have  $\text{area}(\text{circular sector } OPCR) \approx \text{area}(\triangle OPR) + 4/3 \text{area}(\triangle PCR)$ . Since [2],

$$\text{area}(P_n) = 2^{n+1}b_n, \quad b_n = \tan \frac{\pi}{2^{n+1}}, \quad \text{area}(p_n) = 2^{n-1}s_{n-1}, \quad s_n = 2 \sin \frac{\pi}{2^{n+1}},$$

then by elementary calculations (see fig. 2 for notation)

$$\begin{aligned} \pi &\approx 2^{n+1}(\text{area}(\triangle OPR) + 4/3 \cdot \text{area}(\triangle PCR)) \\ &= 2^{n+1} \left( \frac{4}{3} \frac{s_n}{2} \left( 1 - \sqrt{1 - \frac{s_n^2}{4}} \right) \right) + 2^{n-1}s_{n-1} \\ (10) \quad &= 2^{n+1} \cdot \frac{4}{3} \sin \frac{\pi}{2^{n+1}} \left( 1 - \cos \frac{\pi}{2^{n+1}} \right) + 2^n \sin \frac{\pi}{2^n} \\ &= \frac{4}{3} \text{area}(p_{n+1}) - \frac{1}{3} \text{area}(p_n), \end{aligned}$$

which is (5). Let scaling factor  $\alpha$  satisfy  $0 < \alpha < 1$ . Consider now a parabola passing

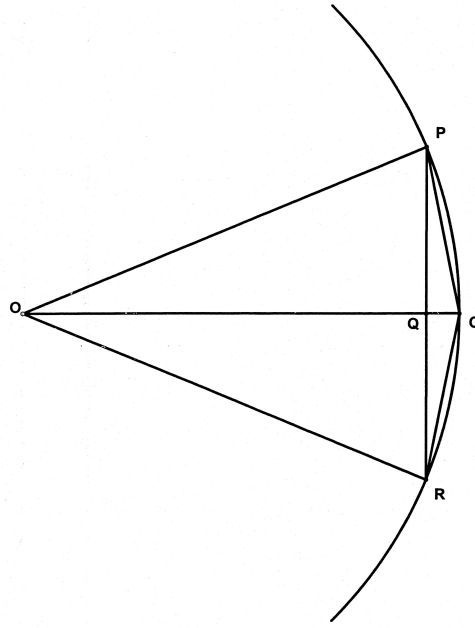
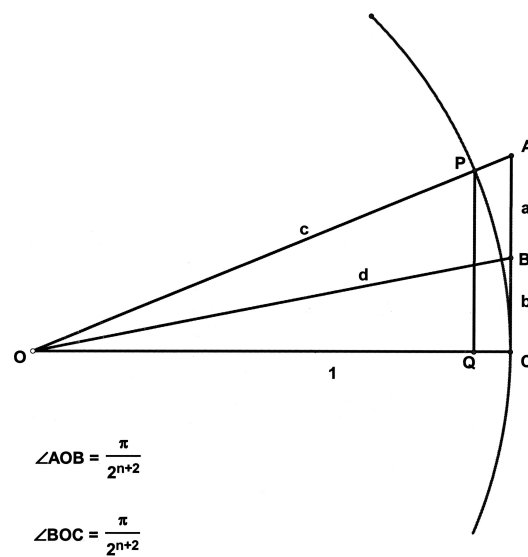


Figure 2:  $|PC| = S_{n+1}$ ,  $|OQ| = h_n$ ,  $|OC| = 1$ ,  $|PR| = S_n$

through points  $A, T, S$  in fig. 3 and the vertex of (convex up) parabola at  $T$  on the (sliding) line segments  $AT, TS$  depending on the value of  $\alpha$  such that  $\alpha = 0, 1$  corresponds with  $T = C, Q$  respectively. Archimedes' result for the area of a parabolic segment and approximation yields that

$$\begin{aligned}
 \pi &\approx 2^{n+1}(\triangle OAS - 4/3 \cdot \triangle ATS) \\
 &= 2^{n+1}b_n - 2^{n+1} \cdot \frac{4}{3} b_n \alpha \left( 1 - \sqrt{1 - \frac{s_n^2}{4}} \right) \\
 (11) \quad &= 2^{n+1} \tan \frac{\pi}{2^{n+1}} \left( 1 - \frac{4}{3} \alpha + \frac{4}{3} \alpha \cos \frac{\pi}{2^{n+1}} \right) \\
 &= \left( 1 - \frac{4}{3} \alpha \right) \text{area}(P_n) + \frac{4}{3} \cdot \alpha \cdot \text{area}(p_{n+1}).
 \end{aligned}$$

Notes: Setting  $\alpha = 1/2$  gives (7). We have so far established a relationship between numerical algorithms with a higher order numerical accuracy than in [2], and, the elimination of terms in the Taylor series for the area of the regular polygons. Next, for interest and completeness,



**Figure 4:**  $C = |OA|$ ,  $a = |AB|$ ,  $b = |BC|$ ,  $d = |OB|$ ,  $|OP| = |OC| = 1$

we list new results from [2]: define real numbers  $a, b, c, d$  as follows

$$\begin{aligned} a &= \tan\left(\frac{\pi}{2^{n+1}}\right) - \tan\left(\frac{\pi}{2^{n+2}}\right) \\ b &= \tan\left(\frac{\pi}{2^{n+2}}\right), \\ c &= \sec\left(\frac{\pi}{2^{n+1}}\right), \\ d &= \sec\left(\frac{\pi}{2^{n+2}}\right). \end{aligned}$$

In [2], (figure 4) we have obtained the following polynomial relations

$$\begin{aligned} b^3 + ab^2 + b - a &= 0, \\ c^3 - c^2 - a^2c - a^2 &= 0, \\ d^6 - (a^2 + 1)d^4 + 4a^2d^2 - 4a^2 &= 0. \end{aligned}$$

### Algorithms and Numerical Results

For convenience, we adopt a coordinate system with the origin at the center of the unit circle and the midpoint of one side of  $p_n$ , the inscribed  $2^{n+1}$ -gon on the positive y-axis. We approximate the circle by the parabola interior to the circle and passing through the three points  $(0,1)$  and  $(\pm a, \sqrt{1-a^2})$ , corresponding to points  $C, R, P$  respec. in fig. 3, where the value of  $a$  depends on the number of sides of the inscribed polygon.

These three points are common to the circle and the parabola. It is well-known (see, for example, [3], p. 524) that two distinct conic sections can intersect in no more than four points. Taking the symmetry of the coordinate system into account, there can be no other intersection points, and thus the parabola remains inside the unit circle for all other values of  $x$  between  $-a$  and  $a$ .

We compute the area of the parabolic wedge bordered by the approximating parabola and the two line segments from the origin to the points  $(\pm a, \sqrt{1-a^2})$  by dividing it into two right triangles and a parabolic segment. The right triangles each have area  $a \cdot \sqrt{1-a^2}$ , which arises from the classical formula for the approximation of  $\pi$  by inscribed  $2^n$ -gons. Denote by  $s_n, n \geq 1$ , the side length of a  $2^{n+1}$ -gon inscribed in the unit circle; then  $2a = s_n$ .

It follows by substitution that the two right triangles have combined area  $A_n$ , where

$$A_n = \frac{s_n}{2} \sqrt{1 - \frac{s_n^2}{4}}.$$

For the parabolic segment, we make use of the result of Archimedes that the area of a parabolic segment is equal to  $\frac{4}{3}$  of the area of the triangle with the same base and height. The area  $a_n$  of the parabolic segment is then

$$a_n = \frac{4}{3} \cdot \frac{1}{2}bh,$$

where  $b$  denotes the base and  $h$  the height of the triangle. We have  $b = s_n$ . The height  $h$  is one leg of a right triangle with other leg  $s_n/2$  and hypotenuse  $s_{n+1}$ . It follows that

$$h = \sqrt{s_{n+1}^2 - \frac{s_n^2}{4}}$$

and we have

$$a_n = \frac{2}{3}s_n \sqrt{s_{n+1}^2 - \frac{s_n^2}{4}}.$$

The area of the unit circle is thus approximated by the sum of  $2^n$  of these parabolic wedges:

$$\pi \approx 2^n \cdot (a_n + A_n) = 2^{n+1} \left( \frac{2}{3}s_n \sqrt{s_{n+1}^2 - \frac{s_n^2}{4}} + \frac{s_n}{2} \sqrt{1 - \frac{s_n^2}{4}} \right).$$

The sequence  $\{s_n\}$  of side lengths is given by

$$s_{n+1} = \sqrt{2 - \sqrt{4 - s_n^2}}$$

with  $s_2 = \sqrt{2}$  (see, for example, [2]). We note further that  $s_n = 2 \sin(\pi/2^{n+1})$ .

Numerical evidence, computed with Mathematica, gives the following sequence of approximations to  $\pi$ . Bold digits are exact.

n	$2^{n+1}$	$s_n$	Approximation to $\pi$	% error
<b>1</b>	4	1.414214	<b>3.1045694997</b>	-1.18
<b>2</b>	8	.765367	<b>3.1391475703</b>	-.078
<b>3</b>	16	.390181	<b>3.1414377167</b>	-.0059
<b>4</b>	32	.196034	<b>3.1415829366</b>	$-3.09 \times 10^{-4}$
<b>5</b>	64	.098135	<b>3.1415920458</b>	$-1.93 \times 10^{-5}$
<b>6</b>	128	.049082	<b>3.1415926156</b>	$-1.21 \times 10^{-6}$
<b>7</b>	256	.024541	<b>3.1415926512</b>	$-7.56 \times 10^{-8}$
<b>8</b>	512	.012272	<b>3.1415926534</b>	$-4.72 \times 10^{-9}$
<b>9</b>	1024	.006136	<b>3.141592653580</b>	$-2.95 \times 10^{-10}$
<b>10</b>	2048	.003068	<b>3.1415926535892</b>	$-1.85 \times 10^{-11}$
<b>11</b>	4096	.001534	<b>3.14159265358976</b>	$-1.15 \times 10^{-12}$
<b>12</b>	8192	.000767	<b>3.141592653589791</b>	$-7.21 \times 10^{-14}$
<b>13</b>	16384	.000383	<b>3.1415926535897931</b>	$-4.51 \times 10^{-15}$

This sequence of approximations to  $\pi$  is, owing to the nature of the approximating parabolas, an underestimate. We seek next to find a corresponding upper bound for  $\pi$  through a sequence of parabolic sectors.

We begin by constructing the tangent lines to the unit circle at the endpoints of the inscribed  $2^{n+1}$ -gon  $p_n$ . The symmetry of the coordinate system means that these lines will intersect at a point  $(0, y)$  on the positive y-axis, and we then consider the parabola passing through the three points  $(0, y)$  and  $(\pm a, \sqrt{1 - a^2})$ , corresponding to points  $M, R, P$  respec. in fig. 3, where the value of  $a$  depends on the number of sides of the inscribed polygon.

Routine calculation of the intersection point reveals that  $y = \frac{1}{\sqrt{1 - a^2}}$ . We calculate the area of the parabolic sector as the sum of two areas involving triangles:

$$A_{\text{sector}} = \frac{4}{3}A_{\text{top triangle}} + A_{\text{interior triangle}}.$$

The top triangle has area  $\frac{1}{2}s_n \left( \frac{1}{\sqrt{1 - a^2}} - \sqrt{1 - a^2} \right)$ , and the interior triangle has area  $\frac{1}{2}s_n \sqrt{1 - a^2}$ . Using the relation  $a = s_n/2$  gives

$$A_{\text{sector}} = s_n \left( \frac{4}{3\sqrt{4 - s_n^2}} - \frac{\sqrt{4 - s_n^2}}{12} \right)$$

and by adding up all  $2^{n+1}$  such sectors, the approximation

$$\pi \approx 2^{n+1} \cdot s_n \left( \frac{4}{3\sqrt{4 - s_n^2}} - \frac{\sqrt{4 - s_n^2}}{12} \right).$$

Numerical computation yields the following overestimates to  $\pi$ :

<b>n</b>	<b><math>2^{n+1}</math></b>	<b><math>s_n</math></b>	<b>Approximation to <math>\pi</math></b>	<b>% error</b>
<b>1</b>	4	1.41421356	4.66666667	48.544614
<b>2</b>	8	0.76536686	<b>3.47546896</b>	10.627613
<b>3</b>	16	0.39018064	<b>3.22297468</b>	2.590470
<b>4</b>	32	0.19603428	<b>3.16181816</b>	.643798
<b>5</b>	64	.09813535	<b>3.14664168</b>	.160716
<b>6</b>	128	.04908246	<b>3.14285445</b>	.040164
<b>7</b>	256	.02454308	<b>3.14190808</b>	.010040
<b>8</b>	512	.01227177	<b>3.14167151</b>	.002510
<b>9</b>	1024	.00613591	<b>3.14161237</b>	.000627
<b>10</b>	2048	.00306796	<b>3.14159758</b>	.000157
<b>11</b>	4096	.00153398	<b>3.14159389</b>	.000039
<b>12</b>	8192	.00076699	<b>3.14159296</b>	.000010

We see that this estimate is not as accurate as our previous underestimate. In an effort to improve the accuracy, we change the vertex of the parabola, to the midpoint of the segment connecting (0,1) to the intersection point  $(0, \frac{1}{\sqrt{1-a^2}})$ . This gives

$$\left(0, \frac{1 + \sqrt{1-a^2}}{2\sqrt{1-a^2}}\right)$$

as the third point on the parabola, which corresponds to point  $K$  in fig. 3. Recalculation using

$$A_{\text{sector}} = \frac{4}{3}A_{\text{top triangle}} + A_{\text{interior triangle}}$$

gives a new formula for the area of a sector as

$$A_{\text{sector}} = \frac{4s_n \left(1 + \sqrt{4 - s_n^2}\right) + s_n^3}{12\sqrt{4 - s_n^2}}$$

and a new estimate for  $\pi$ :

$$\pi \approx 2^{n+1} \cdot \frac{4s_n \left(1 + \sqrt{4 - s_n^2}\right) + s_n^3}{12\sqrt{4 - s_n^2}}.$$

Another set of Mathematica calculations yields the following, more accurate, set of overestimates to  $\pi$ :

<b>n</b>	<b><math>2^{n+1}</math></b>	<b><math>s_n</math></b>	<b>Approximation to <math>\pi</math></b>	<b>% error</b>
<b>1</b>	4	1.41421356	<b>3.88561808</b>	23.6841
<b>2</b>	8	0.76536686	<b>3.30370826</b>	5.27489
<b>3</b>	16	0.39018064	<b>3.18220620</b>	1.29277
<b>4</b>	32	0.19603428	<b>3.15170055</b>	.321744
<b>5</b>	64	.09813535	<b>3.14411686</b>	.080348
<b>6</b>	128	.04908246	<b>3.14222353</b>	.0200816
<b>7</b>	256	.02454308	<b>3.14175036</b>	.0050201
<b>8</b>	512	.01227177	<b>3.14163208</b>	.00125499
<b>9</b>	1024	.00613591	<b>3.141602510</b>	$3.13747 \times 10^{-4}$
<b>10</b>	2048	.00306796	<b>3.14159512</b>	$7.84364 \times 10^{-5}$
<b>11</b>	4096	.00153398	<b>3.14159327</b>	$1.96104 \times 10^{-5}$
<b>12</b>	8192	.00076699	<b>3.141592807</b>	$4.90102 \times 10^{-6}$

Our next attempt at an accurate overestimate considers the concave-up parabola passing through the three points  $(\pm b_n, 1)$  and  $(0, h)$ , which are points  $S, A, T$  respec. in fig. 3, where  $b_n$  represents the side length of  $P_n$ , a  $2^{n+1}$ -gon circumscribed about the unit circle, and  $h$  is to be determined.

This parabolic wedge can also be subdivided, into a large isosceles triangle from which a parabolic segment bounded above by the circumscribed polygon and below by the parabola has been removed. Denoting the area of the wedge by  $A_n$ , we then have

$$A_n = \frac{1}{2}b_n - \frac{2}{3}b_n \cdot (1 - h) = \frac{b_n}{6}(4h - 1).$$

Once again, we have a recursive formula (from [2]) for  $\{b_n\}$ :  $b_2 = 2$  and, for  $n \geq 2$ ,

$$b_{n+1} = \frac{b_n}{1 + \sqrt{1 + \left(\frac{b_n}{2}\right)^2}}.$$

Using this expression for  $b_n$  yields

$$\pi \approx 2^{n+1} \frac{b_n}{6} (4h - 1).$$

It remains to determine the value of  $h$ . Taking  $h$  = the length of the segment connecting the origin to the midpoint of one side of the inscribed  $2^{n+1}$ -gon gives

$$h = \sqrt{1 - \left(\frac{s_n}{2}\right)^2}$$

and thus

$$\pi \approx 2^{n+1} \cdot \frac{b_n}{6} \left( 4\sqrt{1 - \left(\frac{s_n}{2}\right)^2} - 1 \right).$$

Numerical work with Mathematica reveals that this formula is an underestimate of  $\pi$ , and it is less accurate than our first approximation.

<b>n</b>	<b><math>2^{n+1}</math></b>	<b><math>b_n</math></b>	<b>Approximation to <math>\pi</math></b>	<b>% error</b>
<b>1</b>	4	2.000000	2.437903	-22.40
<b>2</b>	8	.828427	2.977387	-5.23
<b>3</b>	16	.397825	<b>3.101061</b>	-1.29
<b>4</b>	32	.196983	<b>3.131490</b>	-.322
<b>5</b>	64	.098254	<b>3.139069</b>	-.080
<b>6</b>	128	.049097	<b>3.140962</b>	-.002
<b>7</b>	256	.024545	<b>3.141435</b>	$-5.02 \times 10^{-3}$
<b>8</b>	512	.012272	<b>3.141553</b>	$-1.25 \times 10^{-3}$
<b>9</b>	1024	.006136	<b>3.141583</b>	$-3.14 \times 10^{-4}$
<b>10</b>	2048	.003068	<b>3.141590</b>	$-7.84 \times 10^{-5}$
<b>11</b>	4096	.001534	<b>3.1415920</b>	$-1.96 \times 10^{-5}$
<b>12</b>	8192	.000767	<b>3.1415925</b>	$-4.90 \times 10^{-6}$
<b>13</b>	16384	.000383	<b>3.14159261</b>	$-1.22 \times 10^{-6}$

Taking  $h = 1$  gives an overestimate, since then the parabola becomes the side of the circumscribing polygon and the  $2^{n+1}$  wedges exactly fill that polygon.

If we split the difference and take  $h$  to be the y-coordinate of the midpoint ( $\alpha = 1/2$ , sect. 1, point T in fig. 3) of the segment joining the midpoint of the side of an inscribed  $2^n$ -gon to the point (0,1), we have

$$h = \sqrt{1 - \left(\frac{s_n}{2}\right)^2} + \frac{1}{2} \left(1 - \sqrt{1 - \left(\frac{s_n}{2}\right)^2}\right)$$

or

$$h = \frac{1}{2} + \frac{1}{2} \sqrt{1 - \left(\frac{s_n}{2}\right)^2}.$$

With this value, our approximation to  $\pi$  becomes

$$\pi \approx 2^{n+1} \cdot \frac{b_n}{6} \left(1 + 2 \sqrt{1 - \left(\frac{s_n}{2}\right)^2}\right),$$

which is revealed by Mathematica to be an accurate overestimate of  $\pi$ , with accuracy comparable to that of our underestimate above.



<b>n</b>	<b><math>2^{n+1}</math></b>	<b>Approximation to <math>\pi</math></b>	<b>% error</b>
<b>1</b>	4	<b>3.218951</b>	2.46
<b>2</b>	8	<b>3.145548</b>	.126
<b>3</b>	16	<b>3.141829</b>	.008
<b>4</b>	32	<b>3.141607</b>	$4.66 \times 10^{-4}$
<b>5</b>	64	<b>3.141594</b>	$2.91 \times 10^{-5}$
<b>6</b>	128	<b>3.1415927</b>	$1.81 \times 10^{-6}$
<b>7</b>	256	<b>3.141592657</b>	$1.13 \times 10^{-7}$
<b>8</b>	512	<b>3.1415926538</b>	$7.09 \times 10^{-9}$
<b>9</b>	1024	<b>3.1415926536</b>	$4.43 \times 10^{-10}$
<b>10</b>	2048	<b>3.14159265359</b>	$2.77 \times 10^{-11}$
<b>11</b>	4096	<b>3.1415926535898</b>	$1.73 \times 10^{-12}$
<b>12</b>	8192	<b>3.141592653589796</b>	$1.08 \times 10^{-13}$
<b>13</b>	16384	<b>3.1415926535897934</b>	$6.76 \times 10^{-15}$

We have examined the sensitivity of the accuracy of this approximation to the value of the constant  $\frac{1}{2}$ . Re-calculation with values running through  $\frac{2^{k-1}-1}{2^k}$  for  $k = 2, 3, \dots, 16$  show less accuracy than  $\frac{1}{2}$ . At the same time, re-calculation with  $\frac{2^{k-1}+1}{2^k}$  for  $k = 2, 3, \dots, 16$  yields a sequence of overestimates to  $\pi$  that are less accurate than that achieved with  $\frac{1}{2}$ .

#### References

- [1] Burton, D. M. *The History of Mathematics*, 4th edition, WCB/McGraw Hill, 1999.
- [2] Grossman, G., *On the numerical approximation to  $\pi$* , Journal of Concrete and Applicable Mathematics **5**, 181-196, 2007.
- [3] Weisstein, E.W., "Conic Section", in *CRC Concise Encyclopedia of Mathematics*, 2nd edition, Chapman & Hall/CRC, 2003.

# ON THE UNIFORM SPECTRUM OF BOUNDED FUNCTIONS AND APPLICATIONS TO DIFFERENTIAL EQUATIONS

NGUYEN VAN MINH, GISELE M. MOPHOU, AND GASTON N'GUÉRÉKATA

**ABSTRACT.** This paper is a survey of the (new) concept of uniform spectrum of bounded functions. We review its properties and relationship with classical concepts of spectra of bounded functions, and some applications to differential equations in Banach spaces.

## 1. INTRODUCTION AND NOTATIONS

This paper is a brief survey of the concept of uniform spectrum of bounded functions introduced in 2004 in the literature by Diagana, Minh and N'Guérékata [4] as an essential tool in the study of the existence and uniqueness of an almost automorphic mild solution to the linear differential equation

$$(1) \quad u'(t) = Au(t) + f(t), \quad t \in \mathbb{R},$$

where  $A$  is a (unbounded) linear operator which generates a holomorphic semigroup of linear operators on a Banach space  $\mathbb{X}$ , and the input  $f$  is almost automorphic (in Bochner's sense). The difficulty in this case is that the classical semigroup theory methods developed by Minh, Naito, Batty and others, do not apply since the existence of a mild solution of this equation is not guaranteed; the main reason is that the group of translations is not necessarily strongly continuous in the case of almost automorphic function. The concept of uniform spectrum of bounded functions was introduced to overcome this obstacle. And it turns out to coincide with the well known Carleman spectrum as we will see below. This yields a much broaden applications field of the concept of the uniform spectrum . Further studies in this direction can be found in [6, 7, 8, 14].

Before proceeding, let us fix our notations. Let  $A$  be a linear operator on a complex Banach space  $\mathbb{X}$ . In what follows,  $D(T)$ ,  $\sigma(A)$  and  $\rho(A)$  will denote the domain, spectrum and the resolvent set of  $A$ , respectively. In particular,  $\sigma_i(A)$  stands for  $\sigma(A) \cap i\mathbb{R}$ . The

---

*Date:* December 28, 2008.

1991 *Mathematics Subject Classification.* Primary: 34G10; Secondary: 43A60.

*Key words and phrases.* Analytic semigroup, almost automorphic solution, uniform spectrum, circular spectrum, sums of commuting operators.

field of complex numbers is denoted by  $\mathbb{C}$ .  $\Re z$  and  $\Im z$  denote the real and imaginary parts of a given complex number  $z$ , respectively. We denote by  $\Gamma$  the unit circle in the complex plane. The notation  $BC(\mathbb{R}, \mathbb{X})$  stands for the space of all  $\mathbb{X}$ -valued bounded and continuous functions on  $\mathbb{R}$ . We will denote by  $BUC(\mathbb{R}, \mathbb{X})$  the subspace of  $BC(\mathbb{R}, \mathbb{X})$  consisting of all uniformly continuous and bounded functions. We set  $BC^1(\mathbb{R}, \mathbb{X}) := \{f \in BC(\mathbb{R}, \mathbb{X}) \mid \exists f' \in BC(\mathbb{R}, \mathbb{X})\}$

Let  $\mathcal{M}$  be a closed subspace of  $BC(\mathbb{R}, \mathbb{X})$ . The operator  $\mathcal{A}_{\mathcal{M}}$  of multiplication by  $A$  is defined on  $D(\mathcal{A}_{\mathcal{M}}) := \{g \in \mathcal{M} : g(t) \in D(A) \ \forall t \in \mathbb{R}, Ag(\cdot) \in \mathcal{M}\}$ , and  $\mathcal{A}_{\mathcal{M}}g := Ag(\cdot)$  for all  $g \in D(\mathcal{A}_{\mathcal{M}})$ .

### 1.1. Almost Automorphic Functions.

**Definition 1.1.** (Bochner) A function  $f \in C(\mathbb{R}, \mathbb{X})$  is said to be *almost automorphic* if for any sequence of real numbers  $(s'_n)$ , there exists a subsequence  $(s_n)$  such that

$$(2) \quad \lim_{m \rightarrow \infty} \lim_{n \rightarrow \infty} f(t + s_n - s_m) = f(t)$$

for any  $t \in \mathbb{R}$ .

The limit in (2) means

$$(3) \quad g(t) = \lim_{n \rightarrow \infty} f(t + s_n)$$

is well-defined for each  $t \in \mathbb{R}$  and

$$(4) \quad f(t) = \lim_{n \rightarrow \infty} g(t - s_n)$$

for each  $t \in \mathbb{R}$ .

*Remark 1.2.* It is clear from the definition above that constant functions and continuous periodic and almost periodic functions are almost automorphic.

*Remark 1.3.* Unlike almost periodic functions, an almost automorphic function may not be uniformly continuous (see examples below)

*Remark 1.4.* Because of pointwise convergence the function  $g$  is measurable but not necessarily continuous. It has been proved that if  $g$  is continuous, then  $f$  is uniformly continuous (cf [13]).

**Example 1.5.** The following are classical examples of almost automorphic functions.

## UNIFORM SPECTRUM OF BOUNDED FUNCTIONS

- (i) (Veech) The function  $f : \mathbb{R} \rightarrow \Gamma$  (the unit circle in  $\mathbb{C}$  defined by

$$f(t) := \frac{2 + e^{it} + e^{i\sqrt{2}t}}{|2 + e^{it} + e^{i\sqrt{2}t}|}$$

is almost automorphic but not almost periodic.

- (ii)(Levitan) The function  $f : \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$f(t) := \sin \frac{1}{2 + \cos t + \cos \sqrt{2}t}$$

is almost automorphic but not almost periodic since it is not uniformly continuous

If the limit in (3) is uniform on any compact subset  $K \subset \mathbb{R}$ , we say that  $f$  is compact almost automorphic.

**Theorem 1.6.** *Assume that  $f$ ,  $f_1$ , and  $f_2$  are almost automorphic and  $\lambda$  is any scalar, then the following hold true.*

- (i)  $\lambda f$  and  $f_1 + f_2$  are almost automorphic,
- (ii)  $f_\tau(t) := f(t + \tau)$ ,  $t \in \mathbb{R}$  is almost automorphic,
- (iii)  $\bar{f}(t) := f(-t)$ ,  $t \in \mathbb{R}$  is almost automorphic,
- (iv) The range  $R_f$  of  $f$  is precompact, so  $f$  is bounded.

*Proof.* See [12, Theorems 2.1.3 and 2.1.4], for proofs. □

**Theorem 1.7.** *If  $\{f_n\}$  is a sequence of almost automorphic  $\mathbb{X}$ -valued functions such that  $f_n \mapsto f$  uniformly on  $\mathbb{R}$ , then  $f$  is almost automorphic.*

*Proof.* see [12, Theorem 2.1.10], for proof. □

**Remark 1.8.** If we equip  $AA(\mathbb{X})$ , the space of almost automorphic functions with the sup norm

$$\|f\|_\infty = \sup_{t \in \mathbb{R}} \|f(t)\|$$

then it turns out to be a Banach space. If we denote  $KAA(\mathbb{X})$ , the space of compact almost automorphic  $\mathbb{X}$ -valued functions, then we have

$$AP(\mathbb{X}) \subset KAA(\mathbb{X}) \subset AA(\mathbb{X}) \subset BC(\mathbb{R}, \mathbb{X})$$

**Theorem 1.9.** *If  $f \in AA(\mathbb{X})$  and its derivative  $f'$  exists and is uniformly continuous on  $\mathbb{R}$ , then  $f' \in AA(\mathbb{X})$ .*

*Proof.* See [12, Theorem 2.4.1] for a detailed proof. □

**Theorem 1.10.** *Let us define  $F : \mathbb{R} \mapsto \mathbb{X}$  by  $F(t) = \int_0^t f(s)ds$  where  $f \in AA(\mathbb{X})$ . Then  $F \in AA(\mathbb{X})$  iff  $R_F = \{F(t) \mid t \in \mathbb{R}\}$  is precompact.*

The monograph [12] presents detailed information about almost automorphic functions along with some applications to differential equations.

## 2. SPECTRUM OF BOUNDED FUNCTIONS

In this Section, we recall briefly the classical theory of spectrum of functions from Harmonic Analysis standpoint.

**2.1. Bohr spectrum.** A natural extension of Fourier exponents of periodic functions to almost periodic functions is the notion of Bohr spectrum.

Consider an almost periodic function  $f : \mathbb{R} \rightarrow \mathbb{X}$ . It is well-known that for every  $\lambda \in \mathbb{R}$ , the average

$$a(f, \lambda) := \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T e^{-i\lambda t} f(t) dt$$

exists and is not zero at most at countably many points  $\lambda$ .

The set  $\sigma_b(f) := \{\lambda \in \mathbb{R} : a(f, \lambda) \neq 0\}$  is called the Bohr spectrum of  $f$ .

**2.2. Carleman spectrum.** Let  $u \in L^\infty(\mathbb{R}, \mathbb{X})$ . The Carleman transform  $\hat{u}$  of  $u$  is defined by

$$\hat{u}(\lambda) := \begin{cases} \int_0^\infty e^{-\lambda t} u(t) dt & (Re \lambda > 0) \\ -\int_0^\infty e^{\lambda t} u(-t) dt & (Re \lambda < 0), \end{cases}$$

Thus  $\hat{u}$  is a holomorphic function defined on  $\mathbb{C} \setminus i\mathbb{R}$ .

**Definition 2.1.** A point  $i\xi \in i\mathbb{R}$  is called regular for  $\hat{u}$  if  $\hat{u}$  has a holomorphic extension to a neighborhood of  $i\xi$ , i.e. there exists an open neighborhood  $U$  of  $i\xi$  and a holomorphic function  $h : U \rightarrow \mathbb{X}$  such that  $h(\lambda) = \hat{u}(\lambda)$  for all  $\lambda \in U \setminus i\mathbb{R}$ .

The Carleman spectrum  $sp_c(u)$  of  $u$  is defined by

$$sp_c(u) := \{\xi \in \mathbb{R} : i\xi \text{ is not regular}\}.$$

For each  $u \in BC(\mathbb{R}, \mathbb{X})$  we let  $\mathcal{M}_u := \overline{\text{span}\{S(\tau)u, \tau \in \mathbb{R}\}}$  (here  $S(\tau)$  denotes the translation  $BC(\mathbb{R}, \mathbb{X}) \ni f(\cdot) \mapsto f(\tau + \cdot) \in BC(\mathbb{R}, \mathbb{X})$ ). It is a closed subspace of  $BC(\mathbb{R}, \mathbb{X})$ .

Note that if

## UNIFORM SPECTRUM OF BOUNDED FUNCTIONS

- $u \in BUC(\mathbb{R}, \mathbb{X})$ , the Carleman spectrum of  $u$  coincides with its Arveson spectrum, defined by (see [2, Lemma 4.6.8])

$$(5) \quad i \operatorname{sp}_c(u) = \sigma(\mathcal{D}_u).$$

where  $\mathcal{D}_u$  is the infinitesimal generator of the restriction of the group of translations  $(S(t)|_{\mathcal{M}_u})_{t \in \mathbb{R}}$  to the closed subspace  $\mathcal{M}_u$ .

- If  $f$  is almost periodic, then  $\operatorname{sp}_c(f) = \overline{\sigma_b(f)}$

*Proof.* See Proposition 1.2 [8] □

Below we list some properties of the Carleman spectra of functions. We refer the reader to [2] for more details and information on other properties of the Carleman spectrum.

**Proposition 2.2.** . Let  $u, u_n, v \in BC(\mathbb{R}, \mathbb{X})$  such that  $\lim_{n \rightarrow \infty} \|u_n - u\| = 0$ . Then

- (i)  $\operatorname{sp}_c(u)$  is closed,
- (ii)  $\operatorname{sp}_c(u + v) \subset \operatorname{sp}_c(u) \cup \operatorname{sp}_c(v)$ ,
- (iii) If  $\operatorname{sp}_c(u) = \emptyset$ , then  $u = 0$ ,
- (iv) If  $\operatorname{sp}_c(u_n) \subset \Lambda, \forall n$ , then  $\operatorname{sp}_c(u) \subset \overline{\Lambda}$ .

**Proposition 2.3.** Let  $u \in L^\infty(\mathbb{R}, \mathbb{X})$ .

- (i) If  $\operatorname{sp}_c(u) = \{0\}$ , then  $u$  is constant.
- (ii) If  $\operatorname{sp}_c(u)$  is finite, then  $u$  is a trigonometric polynomial.

**Theorem 2.4.** Let  $u \in BUC(\mathbb{R}, \mathbb{X})$ . If  $\operatorname{sp}_c(u)$  is discrete, then  $u$  is almost periodic.

### 2.3. Beurling spectrum.

**Definition 2.5.** Let  $u \in L^\infty(\mathbb{R}, \mathbb{X})$ . The Beurling spectrum  $\operatorname{sp}_B(u)$  of  $u$  is the set

$$\operatorname{sp}_B(u) := \{\xi \in \mathbb{R} : \forall \epsilon > 0 \exists v \in L^1(\mathbb{R}) \text{ s.t. } \operatorname{supp}\{v\} \subset (\xi - \epsilon, \xi + \epsilon) \text{ and } u * v \neq 0\}$$

The following relation between the Carleman spectrum and the Beurling spectrum holds.

**Theorem 2.6.** Let  $u \in L^\infty(\mathbb{R}, \mathbb{X})$ . Then

$$\operatorname{sp}_c(u) = \operatorname{sp}_B(u).$$

*Proof.* See [2, Proposition 4.8.4]. □

Now we present a (new) notion of spectrum of bounded functions from Differential Equations standpoint.

3. UNIFORM SPECTRUM OF A FUNCTION IN  $BC(\mathbb{R}, \mathbb{X})$ 

Consider the evolution equation

$$(6) \quad \frac{du(t)}{dt} = Au(t) + f(t),$$

where  $A$  generates a  $C_0$ -semigroup  $(T(t))_{t \geq 0}$  on a Banach space  $\mathbb{X}$ ,  $f$  is an almost periodic function.

**Theorem 3.1.** *If  $(T(t))_{t \geq 0}$  is analytic,  $f$  is almost periodic and*

$$\sigma(A) \cap \overline{i\sigma_F(f)} = \emptyset,$$

*then there is a unique almost periodic mild solution  $u$  such that  $\sigma_b(u) \subset \overline{\sigma_F(f)}$ ; here  $\sigma_F(f)$  denotes the set of frequencies of  $f$ .*

This important result is now a classical one. Our concern is to extend it to the case where  $f$  is almost automorphic. But first, let's make the following remarks.

*Remark 3.2.* In all methods used to study this problem, the uniform continuity of  $f$  plays an important role. For instance that yields:

- The strong continuity of the translation group;

- $\sigma(\mathcal{D}_f) = \overline{i\sigma_F(f)}$

where  $\mathcal{D}_f$  is the differential operator on the closed subspace of  $BC(\mathbb{R}, X)$  generated by  $f(t + \bullet)$ ,  $t \in \mathbb{R}$ .

Since an almost automorphic function may not belong to  $BUC(\mathbb{R}, X)$ , the condition

$$\sigma(A) \cap \text{isp}_c(f) = \emptyset$$

is no longer sufficient for Eq(2) to have a unique mild solution.

To fix this problem, Diagana, Minh and N'Guérékata introduced the concept of *uniform spectrum* as an extension of the Carleman spectrum in their pioneering paper [4] as follows.

Consider the following simple ordinary differential equation in a complex Banach space  $\mathbb{X}$

$$(7) \quad x'(t) - \lambda x = f(t),$$

where  $f \in BC(\mathbb{X})$ . If  $\Re \lambda \neq 0$ , the homogeneous equation associated with this has an exponential dichotomy, so, (7) has a unique bounded solution which we denote by  $x_{f,\lambda}(\cdot)$ .

Moreover, from the theory of ordinary differential equations, it follows that for every fixed  $\xi \in \mathbb{R}$ ,

$$(8) \quad x_{f,\lambda}(\xi) := \begin{cases} \int_{-\infty}^{\xi} e^{\lambda(\xi-t)} f(t) dt & (\text{if } \operatorname{Re} \lambda < 0) \\ -\int_{\xi}^{\infty} e^{\lambda(\xi-t)} f(t) dt & (\text{if } \operatorname{Re} \lambda > 0). \end{cases}$$

$$(9) \quad = \begin{cases} \int_{-\infty}^0 e^{-\lambda\eta} f(\xi + \eta) d\eta & (\text{if } \operatorname{Re} \lambda < 0) \\ -\int_0^{\infty} e^{-\lambda\eta} f(\xi + \eta) d\eta & (\text{if } \operatorname{Re} \lambda > 0). \end{cases}$$

As is well known, the differentiation operator  $\mathcal{D}$  is a closed operator on  $BC(\mathbb{R}, \mathbb{X})$ . The above argument shows that  $\rho(\mathcal{D}) \supset \mathbb{C} \setminus i\mathbb{R}$  and  $x_{f,\lambda} = (\mathcal{D} - \lambda)^{-1}f$  for every  $\lambda \in \mathbb{C} \setminus i\mathbb{R}$  and  $f \in BC(\mathbb{R}, \mathbb{X})$ .

Hence, for every  $\lambda \in \mathbb{C}$  with  $\Re \lambda \neq 0$  and  $f \in BC(\mathbb{R}, \mathbb{X})$  the function  $[(\lambda - \mathcal{D})^{-1}f](t) = \widehat{S(t)f}(\lambda) \in BC(\mathbb{R}, \mathbb{X})$ . Moreover,  $(\lambda - \mathcal{D})^{-1}f$  is analytic on  $\mathbb{C} \setminus i\mathbb{R}$ .

**Definition 3.3.** Let  $f$  be in  $BC(\mathbb{R}, \mathbb{X})$ . Then,

- (i)  $\alpha \in \mathbb{R}$  is said to be *uniformly regular* with respect to  $f$  if there exists a neighborhood  $\mathcal{U}$  of  $i\alpha$  in  $\mathbb{C}$  such that the function  $(\lambda - \mathcal{D})^{-1}f$ , as a complex function of  $\lambda$  with  $\Re \lambda \neq 0$ , has an analytic continuation into  $\mathcal{U}$ .
- (ii) The set of  $\xi \in \mathbb{R}$  such that  $\xi$  is not uniformly regular with respect to  $f \in BC(\mathbb{R}, \mathbb{X})$  is called *uniform spectrum* of  $f$  and is denoted by  $sp_u(f)$ .

If  $f \in BUC(\mathbb{R}, \mathbb{X})$ , then  $\alpha \in \mathbb{R}$  is uniformly regular if and only if it is regular with respect to  $f$ . In fact, this follows from (5) via the identity

$$R(\lambda, \mathcal{D}_f)f = \int_0^{\infty} e^{-\lambda\xi} S(\xi)f d\xi, \quad \Re \lambda \neq 0.$$

(Here the integral is taken in the space  $BUC(\mathbb{R}, \mathbb{X})$ ). For  $f \in BC(\mathbb{R}, \mathbb{X})$ , in general, we do not know if the above (5) holds. We now list some properties of uniform spectra of functions in  $BC(\mathbb{R}, \mathbb{X})$ .

**Proposition 3.4.** Let  $g, f, f_n \in BC(\mathbb{R}, \mathbb{X})$  such that  $f_n \rightarrow f$  as  $n \rightarrow \infty$  and let  $\Lambda \subset \mathbb{R}$  be a closed subset satisfying  $sp_u(f_n) \subset \Lambda$  for all  $n \in \mathbb{N}$ . Then the following assertions hold:

- (i)  $sp_u(f) = sp_u(f(h + \cdot))$ ;
- (ii)  $sp_u(\alpha f(\cdot)) \subset sp_u(f)$ ,  $\alpha \in \mathbb{C}$ ;



- (iii)  $sp_C(f) \subset sp_u(f)$ ;
- (iv)  $sp_u(Bf(\cdot)) \subset sp_u(f)$ ,  $B \in L(\mathbb{X})$ ;
- (v)  $sp_u(f + g) \subset sp_u(f) \cup sp_u(g)$ ;
- (vi)  $sp_u(f) \subset \Lambda$ ;
- (vii) If  $f \in BC^1(\mathbb{R}, \mathbb{X})$ , then  $sp_u(f') \subset sp_u(f)$ .

*Proof.* For the proof see [3] and [4]. □

**Corollary 3.5.** For any closed subset  $\Lambda \subset \mathbb{R}$ , the set  $\Lambda_u(\mathbb{X}) := \{f \in BC(\mathbb{R}, \mathbb{X}) : sp_u(f) \subset \Lambda\}$  is a closed subspace of  $BC(\mathbb{R}, \mathbb{X})$  which is invariant under translations.

*Proof.* Obvious. It suffices to use the properties of uniform spectrum above. □

We also have the following

**Lemma 3.6.** Let  $\Lambda$  be a closed subset of  $\mathbb{R}$  and let  $\mathcal{D}_{\Lambda_u}$  be the differentiation operator acting on  $\Lambda_u(\mathbb{X})$ . Then we have

$$(10) \quad \sigma(\mathcal{D}_{\Lambda_u}) = i\Lambda.$$

*Proof.* Since the function  $g_\alpha$  defined by  $g_\alpha(t) := e^{i\alpha t}x$ ,  $\alpha \in \mathbb{R}, t \in \mathbb{R}, x \neq 0$ , is in  $\Lambda_u(\mathbb{X})$  and  $sp_u(g_\alpha) = sp(g_\alpha) = \{\alpha\}$  we see that  $i\alpha \in \sigma(\mathcal{D}_{\Lambda_u})$ , that is,  $i\Lambda \subset \sigma(\mathcal{D}_{\Lambda_u})$ .

Now let us prove the converse. For  $\beta \in \mathbb{R} \setminus \Lambda$  we consider the equation

$$(11) \quad i\beta g - g' = f, \quad f \in \Lambda_u(\mathbb{X}).$$

We will prove that (11) is uniquely solvable for every  $f \in \Lambda_u(\mathbb{X})$ . This equation has at most one solution. In fact, if  $g_1, g_2$  are two solutions, then  $g = g_1 - g_2$  is a solution of the homogeneous equation, that is for  $f = 0$ . Taking the Carleman transform of both sides of the corresponding equation, we see that  $sp_c(g) \subset \{\beta\}$ . Since  $g \in \Lambda_u(\mathbb{X})$  we have  $sp_c(g) \subset \Lambda$ . Combining these facts we have  $sp_c(g) = \emptyset$ , that is  $g = 0$ .

Now we prove the existence of at least one solution to Eq. (11). In fact, for  $\Re \lambda \neq 0$  Eq. (11) has a unique solution which is nothing but  $(\lambda - \mathcal{D})^{-1}f$ . Since  $i\beta \notin sp_u(f)$ , by definition, the function  $(\lambda - \mathcal{D})^{-1}f$ , defined on  $\mathbb{C} \setminus i\mathbb{R}$ , has an analytic continuation into a neighborhood of  $i\beta$ . In particular, the following limit exists  $\lim_{\lambda \rightarrow i\beta} (\lambda - \mathcal{D})^{-1}f := g_0$ .

## UNIFORM SPECTRUM OF BOUNDED FUNCTIONS

We are going to show that  $g_0$  is a solution of (11) and  $g_0 \in \Lambda_u$ . Indeed, since

$$\begin{aligned} (i\beta - \mathcal{D})(\lambda - \mathcal{D})^{-1}f &= ((i\beta - \lambda) + (\lambda - \mathcal{D}))(\lambda - \mathcal{D})^{-1}f \\ &= (i\beta - \lambda)(\lambda - \mathcal{D})^{-1}f + (\lambda - \mathcal{D})(\lambda - \mathcal{D})^{-1}f \\ &= (i\beta - \lambda)(\lambda - \mathcal{D})^{-1}f + f, \end{aligned}$$

we have

$$(12) \quad \lim_{\lambda \rightarrow i\beta} (i\beta - \mathcal{D})(\lambda - \mathcal{D})^{-1}f = f.$$

Using the closedness of the operator  $(i\beta - \mathcal{D})$ , we come up with  $g_0$  being in the domain of  $i\beta - \mathcal{D}$  and  $(i\beta - \mathcal{D})g_0 = f$ . Next, to show that  $g_0 \in \Lambda_u(\mathbb{X})$ , in view of Corollary 3.5 it suffices to show that for every  $\Re \lambda_0 \neq 0$ , the function  $(\lambda_0 - \mathcal{D})^{-1}f$  is in  $\Lambda_u(\mathbb{X})$ . Since both  $\lambda$  and  $\lambda_0$  are in  $\rho(\mathcal{D})$ , and

$$(\lambda - \mathcal{D})^{-1}(\lambda_0 - \mathcal{D})^{-1}f = (\lambda_0 - \mathcal{D})^{-1}(\lambda - \mathcal{D})^{-1}f$$

we see that  $(\lambda - \mathcal{D})^{-1}(\lambda_0 - \mathcal{D})^{-1}f$  has an analytic continuation into a neighborhood of  $i\beta$ . This completes the proof of the lemma.  $\square$

It turns out that the uniform spectrum of a function  $f \in BC(\mathbb{R}, \mathbb{X})$  coincides with its Carleman spectrum, as shown via this Proposition due to Liu, N'Guérékata, Minh and Vu ([7]):

**Proposition 3.7.** Let  $f \in BC(\mathbb{R}, \mathbb{X})$ . Then

$$(13) \quad sp_u(f) = sp_c(f).$$

*Proof.* First we show that

$$(14) \quad sp_u(x_{f,\lambda_0}) = sp_u(f),$$

where  $x_{f,\lambda}$  is defined by (8), and  $\lambda_0$  is a given complex number such that  $\Re \lambda_0 \neq 0$ . In fact, we have, for every  $\Re \lambda \neq 0$

$$(\lambda - \mathcal{D})^{-1}x_{f,\lambda_0} = -(\lambda - \mathcal{D})^{-1}(\lambda_0 - \mathcal{D})^{-1}f = -(\lambda_0 - \mathcal{D})^{-1}(\lambda - \mathcal{D})^{-1}f.$$

So,  $(\lambda - \mathcal{D})^{-1}x_{f,\lambda_0}$  has an analytic continuation into a neighborhood of  $i\beta$ , where  $\beta$  is a real number, if and only if so does  $(\lambda - \mathcal{D})^{-1}f$ . That is (14) holds. Note that since the derivative of  $x_{f,\lambda_0}$  is bounded, this function is uniformly continuous, so,

$$sp_c(f) \subset sp_u(f) = sp_u(x_{f,\lambda_0}) = sp_c(x_{f,\lambda_0}).$$

To complete the proof of this proposition, it suffices to show that

$$(15) \quad (\mathbb{R} \setminus sp_c(f)) \subset (\mathbb{R} \setminus sp_c(x_{f,\lambda_0})).$$

To this end, we will use the Beurling spectrum as an alternative of the Carleman spectrum (Theorem 3.6). That is,  $\xi \in (\mathbb{R} \setminus sp_c(f))$ , if and only if there is a positive  $\epsilon$  such that if  $\phi \in L^1(\mathbb{R})$  with the support of its Fourier transform  $supp(\tilde{\phi})$  is contained in  $(\xi - \epsilon, \xi + \epsilon)$ , then  $\phi * f = 0$ . Next, it can be easily checked that

$$\phi * x_{f,\lambda_0} = \phi * (\lambda_0 - \mathcal{D})^{-1}f = (\lambda_0 - \mathcal{D})^{-1}\phi * f = 0.$$

This shows that if  $\xi \in (\mathbb{R} \setminus sp_c(f))$ , then  $\xi \in (\mathbb{R} \setminus sp_c(x_{f,\lambda_0}))$ . That is (15) holds. This completes the proof of the proposition.  $\square$

**3.1. Applications to differential equations.** This Section is a slight variation of [4]. We denote by  $\mathcal{F} := \{KAA(\mathbb{X}), AA(\mathbb{X})\}$ . Now let us consider (2) where  $A$  is a (unbounded) linear operator which generates a holomorphic semigroup of linear operators on a Banach space  $\mathbb{X}$  and  $f \in \mathcal{F}$ .

As described earlier, the main difficulty that arises here is concerned with the non-uniform continuity of almost automorphic functions. This implies the non-strong continuity of translation semigroup in the functions space  $\mathcal{F}$ . Therefore, many elegant proofs using semigroup theory fail. To overcome this difficulty we will use the concept of uniform spectrum.

For any closed subset  $\Lambda \subset \mathbb{R}$  we denote

$$\mathcal{F}_\Lambda := \{u \in \mathcal{F} : sp_u(u) \subset \Lambda\}.$$

By the basic properties of uniform spectra of functions,  $\mathcal{F}_\Lambda$  is a closed subspace of  $BC(\mathbb{R}, \mathbb{X})$ . Denote now by  $\mathcal{D}_\Lambda$ , the part of the differential operator  $d/dt$  in  $\mathcal{F}_\Lambda$ . We have the following:

**Lemma 3.8.**

$$(16) \quad \sigma(\mathcal{D}_\Lambda) = i\Lambda.$$

*Proof.* The proof follows the one of Lemma 2.5 [4].  $\square$

Now consider the operator  $\mathcal{A}_\Lambda$  of multiplication by  $A$  and the differential operator  $d/dt$  on the function space  $\mathcal{F}_\Lambda$ .

By definition the operator  $\mathcal{A}_\Lambda$  of multiplication by  $A$  is defined on  $D(\mathcal{A}_\Lambda) := \{g \in \mathcal{F}_\Lambda : g(t) \in D(A) \forall t \in \mathbb{R}, Ag(\cdot) \in \mathcal{F}_\Lambda\}$ , and  $\mathcal{A}g := Ag(\cdot)$  for all  $g \in D(\mathcal{A}_\Lambda)$ . We have the following important result

**Theorem 3.9.** *Assume that  $\Lambda \subset \mathbb{R}$  is closed. Then the operator  $\mathcal{A}_\Lambda$  of multiplication by  $A$  in  $\mathcal{F}_\Lambda$  is the infinitesimal generator of an analytic  $C_0$ -semigroup on  $\mathcal{F}_\Lambda$ .*

*Proof.* We will prove that  $\mathcal{A}_\Lambda$  is a sectorial operator on  $\mathcal{F}_\Lambda$ . In fact, first we check that  $\mathcal{A}_\Lambda$  is densely defined. Consider the semigroup  $\mathcal{T}_\Lambda(t)$  of operators of multiplication by  $T(t)$  on  $\mathcal{F}_\Lambda$ . We now show that it is strongly continuous. Indeed, suppose that  $g \in \mathcal{F}_\Lambda$ , since  $R(g)$  is relatively compact we see that the map  $[0, 1] \times \overline{R(g)} \ni (t, x) \mapsto T(t)x \in \mathbb{X}$  is uniformly continuous. Hence,

$$\sup_{s \in \mathbb{R}} \|T(t)g(s) - g(s)\| \rightarrow 0$$

as  $t \rightarrow 0$ , i.e., the  $\mathcal{T}_\Lambda(t)$  is strongly continuous. By definition,  $g \in D(\mathcal{A}_\Lambda)$  if and only if  $g(s) \in D(A)$ ,  $\forall s \in \mathbb{R}$  and  $Ag(\cdot) \in \mathcal{F}_\Lambda$ . Thus,

$$\frac{T(t)g(s) - g(s)}{t} = \frac{1}{t} \int_0^t T(\xi)Ag(s)d\xi, \quad \forall t \geq 0, s \in \mathbb{R}.$$

Therefore,

$$\lim_{t \rightarrow 0^+} \sup_{s \in \mathbb{R}} \left\| \frac{T(t)g(s) - g(s)}{t} - \frac{1}{t} \int_0^t T(\xi)Ag(s)d\xi \right\| = 0,$$

i.e.,  $g$  is in  $D(G)$ , where  $G$  is the generator of  $\mathcal{T}_\Lambda(t)$  and  $\mathcal{A}_\Lambda g = Gg$ . Conversely, we can easily show that  $G \subset \mathcal{A}_\Lambda$ .

Now it suffices to prove that  $\sigma(\mathcal{A}_\Lambda) \subset \sigma(A)$  to claim that  $\mathcal{A}_\Lambda$  is a sectorial operator. In fact, let  $\mu \in \rho(A)$ . To prove that  $\mu \in \rho(\mathcal{A}_\Lambda)$  we show that for each  $h \in \mathcal{F}_\Lambda$  the equation  $\mu g - \mathcal{A}_\Lambda g = h$  has a unique solution in  $\mathcal{F}_\Lambda$ .

But this follows from the fact that  $(\mu - \mathcal{A}_\Lambda)^{-1}h(\cdot) \in \mathcal{F}_\Lambda$  and that the equation

$$\mu x - Ax = y$$

has a unique solution  $x$  in  $\mathbb{X}$  for any  $y \in \mathbb{X}$ . □

Now observe as in [4] that if  $A$  be the generator of an analytic semigroup, then the operator  $\mathcal{A}_\Lambda$  of multiplication by  $A$  and the differential operator  $\mathcal{D}_\Lambda$  on  $\mathcal{F}_\Lambda$  are commuting, i.e.

$$(17) \quad R(1, \mathcal{D}_\Lambda)\mathcal{T}_\Lambda(\tau) = \mathcal{T}_\Lambda(\tau)R(1, \mathcal{D}_\Lambda), \quad \forall \tau \geq 0.$$

So, by the spectral properties of sums of commuting operators, we have the following result.

**Theorem 3.10.** *If  $\sigma(A) \cap i\Lambda = \emptyset$ , then for every  $f \in \mathcal{F}_\Lambda$  there exists a unique  $u \in \mathcal{F}_\Lambda$  such that*

$$\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda}u = f.$$

*Proof.* Since  $\mathcal{A}_\Lambda$  and  $\mathcal{D}_\Lambda$  commute and satisfy Condition P, the sum  $\mathcal{A}_\Lambda + \mathcal{D}_\Lambda$  is closable (denote its closure by  $\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda}$ ).

Using the fact that  $\sigma(A) \cap i\Lambda = \emptyset$  and properties of commuting operators, we see that  $0 \in \rho(\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda})$ .

Therefore for every  $f \in \mathcal{F}_\Lambda$  there exists a unique  $u \in D(\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda})$  such that

$$\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda}u = f.$$

□

Now we will relate it with the notion of mild solutions to evolution equations.

**Lemma 3.11.** *Let  $u, f \in \mathcal{F}$ . If  $u \in D(\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda})$  and  $\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda}u = f$ , then  $u$  is a mild solution of Eq. (4.11).*

*Proof.* In fact the lemma follows immediately from the following:

For every  $u \in \mathcal{F}$  we say that it belongs to  $D(L)$  of an operator  $L$  acting on  $\mathcal{F}$  if there is a function  $f \in \mathcal{F}$  such that

$$(18) \quad u(t) = T(t-s)u(s) + \int_s^t T(t-\xi)f(\xi)d\xi, \quad \forall t \geq s, t, s \in \mathbb{R}.$$

By a similar argument as in the proof of [10, Lemma 3.1] we can prove that  $L$  is a closed single-valued linear operator acting on  $\mathcal{F}$  which is an extension of  $\mathcal{A}_\Lambda + \mathcal{D}_\Lambda$ .

Thus,  $L$  is an extension of  $\overline{\mathcal{A}_\Lambda + \mathcal{D}_\Lambda}$ . This yields that  $u$  is a mild solution of Eq. (4.11). □

As an immediate consequence of the above argument we have:

**Theorem 3.12.** *Let  $A$  be the generator of an analytic semigroup and let  $\Lambda$  be a closed subset of  $\mathbb{R}$ .*

*Then it is necessary and sufficient for each  $f \in \mathcal{F}_\Lambda$  there exists a unique mild solution  $u \in \mathcal{F}_\Lambda$  to Eq. (4.11) that the condition  $\sigma(A) \cap i\Lambda = \emptyset$  holds.*

*Proof.* The sufficiency follows from the above argument. The necessity can be shown as follows:

For every  $\xi \in \Lambda$ , obviously that the function  $h : \mathbb{R} \ni t \mapsto ae^{i\xi t}$  is in  $\mathcal{F}_\Lambda$ , where  $a \in \mathbb{X}$  is any given element.

By assumption, there is a unique  $g \in D(\mathcal{A}_\Lambda)$  such that  $i\xi g(t) - Ag(t) = h(t)$  for all  $t \in \mathbb{R}$ .

One can easily show that  $g(t)$  is of the form  $be^{i\xi t}$ . Hence,  $b$  is the unique solution of the equation  $i\xi b - Ab = a$ .

That is  $i\xi \notin \sigma(\mathcal{A}_\Lambda)$ , so  $i\Lambda \cap \sigma(\mathcal{A}_\Lambda) = \emptyset$ . □

**Theorem 3.13.** *Let  $A$  be the generator of an analytic semigroup such that  $\sigma(A) \cap i \operatorname{sp}_u(f) = \emptyset$ .*

*Then Eq. (4.11) has a unique mild solution  $w$  in  $\mathcal{F}$  such that  $\operatorname{sp}_u(w) \subset \operatorname{sp}_u(f)$ .*

*Proof.* Set  $\Lambda = \operatorname{sp}_u(f)$ . Then by the above argument we get the theorem. □

*Remark 3.14.* Note that Theorem 4.13 is an extension of Theorem 4.1. Indeed if  $f$  is almost periodic (hence uniformly continuous), then

$$\overline{\sigma_F(f)} = \operatorname{sp}_c(f) = \operatorname{sp}_u(f).$$

A version of this theorem can be stated for bounded solutions that are not necessarily uniformly continuous. See [7].

## REFERENCES

1. W. Arendt, F. Răbiger, A. Sourour, Spectral properties of the operators equations  $AX + XB = Y$ , *Quart. J. Math. Oxford (2)*, **45**(1994), 133-149.
2. W. Arendt, C.J.K. Batty, M. Hieber, F. Neubrander, *Vector-valued Laplace transforms and Cauchy problems*, Monographs in Mathematics, **96**, Birkhuser Verlag, Basel, 2001.
3. J. B. Baillon, J. Blot, G. M. N'Guérékata and D. Pennequin, On  $C^n$ -almost periodic solutions to some nonautonomous differential equations in Banach spaces, *Annales Societatis Mathematicae Polonae, Serie 1*, XLVI (2), 263-273, (2006).
4. T. Diagana, G. N'Guérékata, Nguyen Van Minh, Almost automorphic solutions of evolution equations. *Proc. Amer. Math. Soc.* **132** No.11 (2004), 3289-3298.
5. Y. Hino, T. Naito, N.V. Minh, J.S. Shin, *Almost Periodic Solutions of Differential Equations in Banach Spaces*. Taylor & Francis, London - New York, 2002.
6. J.H. Liu, G.M. N'Guérékata, Nguyen van Minh, A Massera type theorem for almost automorphic solutions of differential equations. *J. Math. Anal. Appl.* **299** (2004), no. 2, 587-599.

7. J. Liu, G.M. N'Guérékata, Nguyen van Minh and Q. P. Vu, *Bounded solutions of parabolic equations in continuous function spaces*, Funkcial. Ekvac. **49** (2006), no. 3, 337-355.
8. J. H. Liu, G. M. N'Guérékata, Nguyen van Minh, *Topics on Stability and Periodicity in Abstract Differential Equations*, Series on Concrete and Applicable Mathematics, Vol.6, World Scientific, New Jersey-London-Singapore-Beijing-Shanghai-Hong Kong-Taipei-Chennai, 2008.
9. A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*, Birhauser, Basel, 1995.
10. S. Murakami, T. Naito, N.V. Minh, Evolution semigroups and sums of commuting operators: a new approach to the admissibility theory of function spaces, *J. Differential Equations* **164** (2000), 240-285.
11. T. Naito, Nguyen Van Minh, J. Liu, On the bounded solutions of Volterra equations, *Applicable Analysis*, **83** (2004), 433-446.
12. G. M. N'Guérékata, *Almost Automorphic and Almost Periodic Functions in Abstract Spaces*, Kluwer, Amsterdam, 2001.
13. G. M. N'Guérékata, *Comments on almost automorphic and almost periodic functions in Banach spaces*, Far East J. Math. Sci. (FJMS) **17** (2005), No.3, 337-344.
14. G. M. N'Guérékata, *Almost automorphic solutions to second-order semilinear evolution equations*, Nonlinear Analysis, T.M.A., (in press).
15. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Applied Math. Sci. 44, Spriger-Verlag, Berlin-New York 1983.
16. J. Prüss, *Evolutionary Integral Equations and Applications*, Birkhäuser, Basel, 1993.
17. J. Prüss, Bounded solutions of Volterra equations, *SIAM Math. Anal.* **19**(1987), 133-149.
18. E. Schuler, Vu Quoc Phong, The operator equation  $AX - X\mathcal{D}^2 = -\delta_0$  and second order differential equations in Banach spaces. Semigroups of operators: theory and applications (Newport Beach, CA, 1998), 352-363.
19. E. Sinestrari, On the abstract Cauchy problem of parabolic type in spaces of continuous functions, *J. Math. Anal. Appl.* **107**(1985), 16-66.
20. B. Stewart, Generation of analytic semigroups by strongly elliptic operators. *Trans. Amer. Math. Soc.* **199** (1974), 141-162.
21. Q.P. Vu, E. Schüler, The operator equation  $AX - XB = C$ , stability and asymptotic behaviour of differential equations, *J. Differential Equations* **145** (1998), 394-419.
22. M. Yamaguchi, Existence of periodic solutions of second order nonlinear evolution equations and applications. *Funkc. Ekv.* **38** (1995), 519-538.

## UNIFORM SPECTRUM OF BOUNDED FUNCTIONS

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WEST GEORGIA, CARROLLTON. GA 30118

*E-mail address:* `nguyenvm@jmu.edu`

GISÈLE M. MOPHOU, UNIVERSITÉ DES ANTILLES ET DE LA GUADELOUPE, DÉPARTEMENT DE MATHÉMATIQUES  
ET INFORMATIQUE, UNIVERSITÉ DES ANTILLES ET DE LA GUYANE, CAMPUS FOUILLOLE 97159  
POINTE-À-PITRE GUADELOUPE (FWI)

*E-mail address:* `gmophou@univ-ag.fr`

DEPARTMENT OF MATHEMATICS, MORGAN STATE UNIVERSITY, 1700 E. COLD SPRING LANE,  
BALTIMORE, MD 21251, USA

*E-mail address:* `gnguerrek@jewel.morgan.edu`



# Sequential Decision Model for Exponential Pattern Recognition

Iuliana Florentina Iatan,

Department of Mathematics and Informatics,  
Technical University of Civil Engineering, Bucharest, Romania,  
e-mail iuliafi@yahoo.com

## Abstract

Sequential procedures are concerned with statistical analysis of data when the number of observations is not predetermined.

The purposes of the present paper are [4]:

1. the description of the SPRT (SPRT= Sequential Probability Ratio Test) in the case of exponential patterns;
2. the evaluation of the mean number of samples for the proposed test in our case;
3. the obtaining of signal to noise ratio which defines a measure of two classes separability for the patterns with exponential distribution).

**AMS Subject Classification:** 62-xx, 62Lxx, 62L12.

**Keywords:** Sequential Probability Ratio Test, sequential testing, statistics hypothesis, sequential procedures, sequential classification, exponential patterns, samples, signal to noise ratio.

## 1 Introduction

### 1.1 SPRT for Exponential Type Densities

Let observable random variables  $Y_1, Y_2, \dots$  which are independent and identically distributed, belonging to  $\omega_1$  or  $\omega_2$  class.

We want to test the hypotheses

$$H_0 : X \in \omega_1 \text{ against } H_1 : X \in \omega_2.$$

The probability density functions for the two classes are of exponential type:

$$f(X; \theta_0) = e^{C(X)N(\theta_0)+D(X)+M(\theta_0)}, \text{ for } C(X) < 0$$

in the case of first class and

$$f(X; \theta_1) = e^{C(X)N(\theta_1)+D(X)+M(\theta_1)}, \text{ for } C(X) < 0$$

in the case of second class, where the functions  $N(\theta_i)$ ,  $M(\theta_i)$  satisfy the condition

$$\int f(X; \theta_i) dX = 1, \quad i = \overline{0, 1}.$$

Suppose we observe at first  $X_1$ .

For the first stage we shall have the likelihood ratio:

$$z_1 = \ln \frac{f(X_1; \theta_1)}{f(X_1; \theta_0)} = \ln \frac{e^{C(X_1)N(\theta_1)+D(X_1)+M(\theta_1)}}{e^{C(X_1)N(\theta_0)+D(X_1)+M(\theta_0)}};$$

therefore

$$z_1 = C(X_1)[N(\theta_1) - N(\theta_0)] + M(\theta_1) - M(\theta_0).$$

For the stopping bounds  $0 < B < A < \infty$  we define the sequential probability ratio test at the first stage as:

i) accept  $H_0$  if  $z_1 \leq B$ , namely

$$C(X_1)[N(\theta_1) - N(\theta_0)] + M(\theta_1) - M(\theta_0) \leq B$$

or

$$C(X_1) \leq [N(\theta_1) - N(\theta_0)]^{-1}[B - M(\theta_1) + M(\theta_0)];$$

ii) reject  $H_0$  if  $z_1 \geq A$ , namely

$$C(X_1)[N(\theta_1) - N(\theta_0)] + M(\theta_1) - M(\theta_0) \geq A$$

or

$$C(X_1) \geq [N(\theta_1) - N(\theta_0)]^{-1}[A - M(\theta_1) + M(\theta_0)];$$

iii) continue sampling with  $X_2$  if  $B < z_1 < A$ , therefore

$$B < C(X_1)[N(\theta_1) - N(\theta_0)] + M(\theta_1) - M(\theta_0) < A$$

or

$$U < C(X_1) < V,$$

where

$$U = [N(\theta_1) - N(\theta_0)]^{-1}[B - M(\theta_1) + M(\theta_0)],$$

$$V = [N(\theta_1) - N(\theta_0)]^{-1}[A - M(\theta_1) + M(\theta_0)].$$

The likelihood ratio at the  $n$  stage is

$$Z_n = \sum_{i=1}^n z_i = n[M(\theta_1) - M(\theta_0)] + [N(\theta_1) - N(\theta_0)] \sum_{i=1}^n C(X_i) \quad (1)$$

and the sequential probability ratio test becomes:

i) accept  $H_0$  if

$$\sum_{i=1}^n C(X_i) \leq [N(\theta_1) - N(\theta_0)]^{-1}[B - n(M(\theta_1) - M(\theta_0))];$$

ii) reject  $H_0$  if

$$\sum_{i=1}^n C(X_i) \geq [N(\theta_1) - N(\theta_0)]^{-1}[A - n(M(\theta_1) - M(\theta_0))];$$

iii) continue sampling with  $X_{n+1}$  if

$$W < \sum_{i=1}^n C(X_i) < T,$$

where

$$W = [N(\theta_1) - N(\theta_0)]^{-1}[B - n(M(\theta_1) - M(\theta_0))],$$

$$T = [N(\theta_1) - N(\theta_0)]^{-1}[A - n(M(\theta_1) - M(\theta_0))].$$

## 2 Description of Sequential Probability Ratio Test

### 2.1 SPRT in case of the pattern having an exponential distribution

Consider the following probability density functions for the two classes:

$$f(X; \lambda_0) = \begin{cases} \lambda_0 \cdot e^{-\lambda_0 X}, & \text{for } X \geq 0, \\ 0, & \text{otherwise} \end{cases}$$

in the case of first class and

$$f(X; \lambda_1) = \begin{cases} \lambda_1 \cdot e^{-\lambda_1 X}, & \text{for } X \geq 0, \\ 0, & \text{otherwise} \end{cases}$$

in the case of second class, with  $\lambda_0, \lambda_1 > 0$  and  $\lambda_1 > \lambda_0$ .

Suppose we observe at first  $X_1$ .

We shall calculate

$$z_1 = \ln \frac{f(X_1; \lambda_1)}{f(X_1; \lambda_0)} = \ln \frac{\lambda_1 \cdot e^{-\lambda_1 X_1}}{\lambda_0 \cdot e^{-\lambda_0 X_1}} = \ln \frac{\lambda_1}{\lambda_0} - X_1(\lambda_1 - \lambda_0). \quad (2)$$

In our case we define the sequential probability ratio test at the first stage as:

i) accept  $H_0$  if  $z_1 \leq B$ , namely

$$\ln \frac{\lambda_1}{\lambda_0} - X_1(\lambda_1 - \lambda_0) \leq B$$

or

$$(\lambda_0 - \lambda_1)^{-1} \left[ B + \ln \frac{\lambda_0}{\lambda_1} \right] \leq X_1$$

ii) reject  $H_0$  dacă  $z_1 \geq A$ , namely

$$\ln \frac{\lambda_1}{\lambda_0} - X_1(\lambda_1 - \lambda_0) \geq A$$

or

$$X_1 \leq (\lambda_0 - \lambda_1)^{-1} \left[ A + \ln \frac{\lambda_0}{\lambda_1} \right]$$

iii) continue sampling with  $X_2$  if  $B < z_1 < A$ , therefore

$$B < \ln \frac{\lambda_1}{\lambda_0} - X_1(\lambda_1 - \lambda_0) < A$$

or

$$(\lambda_0 - \lambda_1)^{-1} \left[ B + \ln \frac{\lambda_0}{\lambda_1} \right] \leq X_1 \leq (\lambda_0 - \lambda_1)^{-1} \left[ A + \ln \frac{\lambda_0}{\lambda_1} \right].$$

The likelihood ratio at the  $n$  stage is

$$Z_n = \sum_{i=1}^n z_i = n \ln \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0) \sum_{i=1}^n X_i \quad (3)$$

and the sequential probability ratio test becomes:

i) accept  $H_0$  if

$$(\lambda_0 - \lambda_1)^{-1} \left[ B + n \ln \frac{\lambda_0}{\lambda_1} \right] \leq \sum_{i=1}^n X_i$$

ii) reject  $H_0$  if

$$\sum_{i=1}^n X_i \leq (\lambda_0 - \lambda_1)^{-1} \left[ A + n \ln \frac{\lambda_0}{\lambda_1} \right]$$

iii) continue sampling with  $X_{n+1}$  if

$$(\lambda_0 - \lambda_1)^{-1} \left[ B + n \ln \frac{\lambda_0}{\lambda_1} \right] \leq \sum_{i=1}^n X_i \leq (\lambda_0 - \lambda_1)^{-1} \left[ A + n \ln \frac{\lambda_0}{\lambda_1} \right].$$

**Theorem 1** *The selection mean volume of sequential probability ratio test is given by the formula*

$$E[n] = p_0 \frac{B(1 - \alpha) + A\alpha}{\ln \frac{\lambda_1}{\lambda_0} - \left( \frac{\lambda_1}{\lambda_0} - 1 \right)} + p_1 \frac{B\beta + A(1 - \beta)}{\ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_0}{\lambda_1} - 1 \right)}, \quad (4)$$

- $p_0$  and  $p_1$  are the probabilities that the hypothesis  $H_0$  and respectively  $H_1$  be trues;

- $A$  and  $B$  are the stopping bounds for the sequential probability ratio test ;
- $\alpha$  and  $\beta$  are the error probabilities of first and second kind.

**Proof:**

We need the following three stages in order to determine the selection mean volume:

1. Calculate the conditional mean of  $z_1$  given  $\lambda_i$ ,  $i = \overline{0, 1}$ :  $E_{\lambda_i}[z_1]$ .
  2. Calculate the conditional mean of  $n$  given  $\lambda_i$ ,  $i = \overline{0, 1}$ :  $E_{\lambda_i}[n]$ .
  3. Calculate  $E[n]$ - the selection mean volume of the sequential probability ratio test.
- Stage 1. We shall have

$$E_{\lambda_0}[z_1] = E_{\lambda_0} \left[ \ln \frac{\lambda_1}{\lambda_0} - X_1(\lambda_1 - \lambda_0) \right] = \ln \frac{\lambda_1}{\lambda_0} - (\lambda_1 - \lambda_0)E_{\lambda_0}[X_1];$$

therefore

$$E_{\lambda_0}[z_1] = \ln \frac{\lambda_1}{\lambda_0} - \frac{\lambda_1 - \lambda_0}{\lambda_0} = \ln \frac{\lambda_1}{\lambda_0} - \left( \frac{\lambda_1}{\lambda_0} - 1 \right). \quad (5)$$

Similarly, we obtain

$$E_{\lambda_1}[z_1] = \ln \frac{\lambda_1}{\lambda_0} - \frac{\lambda_1 - \lambda_0}{\lambda_1} = \ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_0}{\lambda_1} - 1 \right). \quad (6)$$

- Stage 2. Because of fact that

$$E_{\lambda_i}[n] = \frac{B(1 - \alpha) + A\alpha}{E_{\lambda_i}[z_1]}, \quad i = \overline{0, 1} \quad (7)$$

we have

$$E_{\lambda_0}[n] = \frac{B(1 - \alpha) + A\alpha}{\ln \frac{\lambda_1}{\lambda_0} - \left( \frac{\lambda_1}{\lambda_0} - 1 \right)} \quad (8)$$

and

$$E_{\lambda_1}[n] = \frac{B\beta + A(1 - \beta)}{\ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_0}{\lambda_1} - 1 \right)}. \quad (9)$$

- Stage 3. We know that we can determine the selection mean volume of sequential probability ratio test using the formula

$$E[n] = p_0 E_{\lambda_0}[n] + p_1 E_{\lambda_1}[n]; \quad (10)$$

therefore one deduces that

$$E[n] = p_0 \frac{B(1-\alpha) + A\alpha}{\ln \frac{\lambda_1}{\lambda_0} - \left(\frac{\lambda_1}{\lambda_0} - 1\right)} + p_1 \frac{B\beta + A(1-\beta)}{\ln \frac{\lambda_1}{\lambda_0} + \left(\frac{\lambda_0}{\lambda_1} - 1\right)}. \quad (11)$$

■

**Theorem 2** *The signal to noise ratio which defines a measure of classes separability in the case of pattern with exponential distribution, for  $n$  samplings has the expression*

$$d_n^2 = \frac{n \frac{(\lambda_0 - \lambda_1)^4}{\lambda_0^2 \lambda_1^2}}{\ln \frac{\lambda_1}{\lambda_0} - \frac{\lambda_1 - \lambda_0}{\lambda_0^2}}. \quad (12)$$

**Proof:**

From the relation (3) we can note that

1. the conditional mean of  $Z_n$  given  $\lambda_0$  is

$$E_{\lambda_0}[Z_n] = n E_{\lambda_0}[z_1] = n \left[ \ln \frac{\lambda_1}{\lambda_0} - \left( \frac{\lambda_1}{\lambda_0} - 1 \right) \right]; \quad (13)$$

2. the conditional mean of  $Z_n$  given  $\lambda_1$  is

$$E_{\lambda_1}[Z_n] = n E_{\lambda_1}[z_1] = n \left[ \ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_0}{\lambda_1} - 1 \right) \right]; \quad (14)$$

3. the conditional variance of  $Z_n$  given  $\lambda_0$  is

$$\text{Var}_{\lambda_0}[Z_n] = n \cdot \text{Var}_{\lambda_0}[z_1] = n \left[ \ln \frac{\lambda_1}{\lambda_0} - \frac{\lambda_1 - \lambda_0}{\lambda_0^2} \right]; \quad (15)$$

We know [7] that

$$d_n^2 = \frac{\{E_{\lambda_1}[Z_n] - E_{\lambda_0}[Z_n]\}^2}{\text{Var}_{\lambda_0}[Z_n]}. \quad (16)$$

Substituting (13), (14) and (15) into (16) we deduce:

$$d_n^2 = \frac{n^2 T}{n \left[ \ln \frac{\lambda_1}{\lambda_0} - \frac{\lambda_1 - \lambda_0}{\lambda_0^2} \right]}, \quad (17)$$

where

$$\begin{aligned} T &= \left[ \ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_0}{\lambda_1} - 1 \right) \right]^2 - 2 \left[ \ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_0}{\lambda_1} - 1 \right) \right] \left[ \ln \frac{\lambda_1}{\lambda_0} - \left( \frac{\lambda_1}{\lambda_0} - 1 \right) \right] + \\ &+ \left[ \ln \frac{\lambda_1}{\lambda_0} - \left( \frac{\lambda_1}{\lambda_0} - 1 \right) \right]^2 = \left( \ln \frac{\lambda_1}{\lambda_0} \right)^2 + 2 \left( \frac{\lambda_0}{\lambda_1} - 1 \right) \ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_0}{\lambda_1} - 1 \right)^2 - \\ &- 2 \left( \ln \frac{\lambda_1}{\lambda_0} \right)^2 - 2 \left( \frac{\lambda_0}{\lambda_1} - 1 \right) \ln \frac{\lambda_1}{\lambda_0} + 2 \left( \frac{\lambda_1}{\lambda_0} - 1 \right) \ln \frac{\lambda_1}{\lambda_0} + 2 \left( \frac{\lambda_1}{\lambda_0} - 1 \right) \left( \frac{\lambda_0}{\lambda_1} - 1 \right) + \\ &+ \left( \ln \frac{\lambda_1}{\lambda_0} \right)^2 - 2 \left( \frac{\lambda_1}{\lambda_0} - 1 \right) \ln \frac{\lambda_1}{\lambda_0} + \left( \frac{\lambda_1}{\lambda_0} - 1 \right)^2 = \left( \frac{\lambda_0}{\lambda_1} - 1 \right)^2 + \\ &+ 2 \left( \frac{\lambda_1}{\lambda_0} - 1 \right) \left( \frac{\lambda_0}{\lambda_1} - 1 \right) + \left( \frac{\lambda_1}{\lambda_0} - 1 \right)^2 = \left[ \left( \frac{\lambda_1}{\lambda_0} - 1 \right) + \left( \frac{\lambda_0}{\lambda_1} - 1 \right) \right]^2 = \\ &= \left( \frac{\lambda_0}{\lambda_1} + \frac{\lambda_1}{\lambda_0} - 2 \right)^2 = \left( \frac{\lambda_0^2 - 2\lambda_0\lambda_1 + \lambda_1^2}{\lambda_0\lambda_1} \right)^2; \end{aligned}$$

therefore

$$T = \frac{(\lambda_0 - \lambda_1)^4}{\lambda_0^2 \lambda_1^2}. \quad (18)$$

Substituting (18) into (17) we obtain

$$d_n^2 = \frac{n \frac{(\lambda_0 - \lambda_1)^4}{\lambda_0^2 \lambda_1^2}}{\ln \frac{\lambda_1}{\lambda_0} - \frac{\lambda_1 - \lambda_0}{\lambda_0^2}}. \quad (19)$$

■

### 3 Conclusions

- In this paper, we describe in detail the sequential probability ratio test in the case of patterns with an exponential distribution.
- We construct two theorems and we proof each of them.



- Theorem 1 supplies the selection mean volume of the sequential probability ratio test for our considering case.
- Theorem 2 calculates the signal to noise ratio in the case of pattern with exponential distribution, for  $n$  samplings.
- The formula (19) indicates that the more samplings the better separation will be.

## References

- [1] Bishop, C.,M., 2006. *Pattern Recognition and Machine Learning*. Springer, Heidelberg.
- [2] Castelli, V., Cover, T., M., 1995. "On the exponential value of labeled samples". *Pattern Recognition Letters*, 16: 105-111.
- [3] Govindarajulu, Z., 1975. *Sequential Statistical Procedures*. Academic Press, New York.
- [4] **Iatan, I.**, 2008. *Statistical Methods for Pattern Recognition*, second Ph.D. Thesis, Faculty of Mathematics and Computer Science, University of Bucharest, Ph.D. Supervisor: Prof. Dr. Ion Văduva.
- [5] Mihoc, Gh., Craiu, V., 1977. *Treatise of Mathematical Statistics. Testing Statistical Hypotheses*. Vol. II, Ed. Academy of Bucharest.
- [6] Văduva, I., 1962. *Sequential Tests for Exponential Families*(Russian). Rev. Roum. Math. Pures et Appl., Tom VII, 4: 706-716.
- [7] Young, T., Y., Calvert, T., W., 1974. *Classification, Estimation and Pattern Recognition*. Elsevier.

# Inequalities for polynomials with curved majorants dependent on Chebyshev polynomials

Sergey Ivanovich Kalmykov

Institute of Applied Mathematics of

Far Eastern Branch of

the Russian Academy of Sciences, Vladivostok

sergeykalmykov@inbox.ru

March 6, 2009

**ABSTRACT.** Using methods of geometric function theory, we get new inequalities for polynomials with majorants dependent on Chebyshev polynomials of the first kind. These theorems refine some known results for algebraic polynomials with constraints on the the interval.

**Keywords:** Bernstein-type inequality, Chebyshev polynomials, geometric function theory.

## §1. INTRODUCTION

Many inequalities for polynomials with different curved majorants on subsets of complex plane are available today ([1]-[6]). For example, Lachance studied polynomials with the restriction  $(1-x^2)^{\frac{\lambda}{2}}|p(x)| \leq 1$  on the interval  $[-1, 1]$  where  $\lambda$  is a fixed positive integer [5]. In this paper we consider polynomials with

majorants dependent on Chebyshev polynomials of the first kind. Some special cases of our statements were proved earlier in [7], [8]. To obtain estimates of coefficients, covering theorems and Bernstein-type inequalities for polynomials we use an approach proposed by Dubinin V.N. in [6]. This approach consists in constructing an analytic function associated with the given polynomial and applying some methods of geometric function theory to this function. Extremal polynomials in our statements are superpositions of Chebyshev polynomials of different kinds.

Following [9], we introduce the functions

$$\begin{aligned}\mathrm{Co}_\nu(z) &= \frac{1}{2}(h^{2\nu}(z) + h^{-2\nu}(z)), \\ \mathrm{Si}_\nu(z) &= \frac{1}{2}(h^{2\nu}(z) - h^{-2\nu}(z)),\end{aligned}$$

where  $h(z) = \sqrt{(z+1)/2} + \sqrt{(z-1)/2}$ , and  $\nu$  is integer or half-integer. Chebyshev polynomials of the first, second, third and fourth kinds are polynomials

$$T_n(z) = \mathrm{Co}_n(z),$$

$$U_n(z) = \mathrm{Si}_{n+1}(z)/\sqrt{z^2-1},$$

$$V_n(z) = \mathrm{Co}_{n+1/2}(z)/\sqrt{(z+1)/2},$$

$$W_n(z) = \mathrm{Si}_{n+1/2}(z)/\sqrt{(z-1)/2},$$

$n = 1, 2, \dots$  correspondingly. The following representations are well known [9]:

$$T_n(z) = \frac{1}{2}((z + \sqrt{z^2-1})^n + (z - \sqrt{z^2-1})^n), \quad (1)$$

$$U_n(z) = \frac{(z + \sqrt{z^2-1})^{n+1} - (z - \sqrt{z^2-1})^{n+1}}{2\sqrt{z^2-1}}, \quad (2)$$

$$V_n(z) = \frac{(z + \sqrt{z^2-1})^{n+\frac{1}{2}} + (z - \sqrt{z^2-1})^{n+\frac{1}{2}}}{(z + \sqrt{z^2-1})^{\frac{1}{2}} + (z - \sqrt{z^2-1})^{\frac{1}{2}}}, \quad (3)$$

$$W_n(z) = \frac{(z + \sqrt{z^2-1})^{n+\frac{1}{2}} - (z - \sqrt{z^2-1})^{n+\frac{1}{2}}}{(z + \sqrt{z^2-1})^{\frac{1}{2}} - (z - \sqrt{z^2-1})^{\frac{1}{2}}}. \quad (4)$$

In the first section of this paper we give some necessary information from the theory of bounded univalent functions and a modified form of sufficient condition of univalence from the article [6]. In the subsequent sections we obtain exact estimates of coefficients, covering theorems and inequalities for derivatives for polynomials of the form

$$P_n(z) = c_n z^n + \dots + c_0, \quad c_n \neq 0, \quad c_l \in \mathbb{R}, \quad l = 0, 1, \dots, n, \quad n \geq 1, \quad (5)$$

which satisfy one of the following conditions for some positive integer  $k$ :

$$|P(z)| \leq 1/\sqrt{1 - T_k^2(z)}, \quad z \in [-1, 1], \quad (6)$$

$$|P(z)| \leq \sqrt{2/(1 + T_k(z))}, \quad z \in [-1, 1], \quad (7)$$

$$|P(z)| \leq \sqrt{2/(1 - T_k(z))}, \quad z \in [-1, 1]. \quad (8)$$

Let  $\mathcal{PU}_{n,k}$ ,  $\mathcal{PV}_{n,k}$ ,  $\mathcal{PW}_{n,k}$  denote classes of polynomials (5) satisfying conditions (6), (7) and (8), respectively.

## §2. AUXILIARY STATEMENTS

Let a function  $w = f(z)$  be regular and univalent in the unit disk  $U = \{z : |z| < 1\}$ , and let the following conditions be fulfilled:  $f(0) = 0$  and  $|f(z)| < 1$  for  $|z| < 1$ . Denote the class of such functions by  $\mathcal{B}$ . A function  $f \in \mathcal{B}$  has the expansion

$$w = f(z) = \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 + \dots$$

The following inequalities for functions from  $\mathcal{B}$  are well known ([6], [10]):

$$|\alpha_2/\alpha_1| \leq 2(1 - |\alpha_1|), \quad (9)$$

$$\left( \frac{1 + |f(z)|}{1 + |z|} \right)^2 \leq \left| \frac{f(z)}{\alpha_1 z} \right| \leq \left( \frac{1 - |f(z)|}{1 - |z|} \right)^2 \quad \text{for all } z \in U, \quad (10)$$

$$|f'(z)| \geq \sqrt{1/|\alpha_1|}, \quad \text{if } |f(z)| = |z| = 1, \quad (11)$$

and the derivative  $f'(z)$  exists at the boundary point. The equality in (9)-(11) occurs for the function  $f(z) = \alpha_1 z$  with  $|\alpha_1| = 1$ .

We will use these estimates for functions  $f \in \mathcal{B}$  associated with the given polynomial  $P(z)$ . We need the following lemma

**Lemma 1.** *Let a function  $z = F(w)$  be continuous on an open set  $B$ ,  $0 \in B \subset \mathbb{C}_w$ , regular on the set  $B \setminus \{w : |F(w)| = 1\}$  which is a family of domains  $\mathcal{D}$ . Suppose that*

$$\lim_{\substack{w \rightarrow w_0 \\ w \in B}} |F(w)| \geq 1 \text{ for } w_0 \in \partial B.$$

*Further assume that  $F'(0) \neq 0$  and that  $F(w) = 0$  if and only if  $w = 0$ .*

*If  $0 \in D \in \mathcal{D}$  then the function  $F(w)$  effects conformal and univalent mapping of the domain  $D$  onto the unit disk  $U$ . If  $0 \notin D \in \mathcal{D}$  then  $F(D) \subset \{z : |z| > 1\}$ .*

Denote by

$$\zeta = \Phi(\omega) = \omega - \sqrt{\omega^2 - 1}$$

we denote the branch of the analytic function (the inverse of the Zhukovskii mapping) which maps the exterior of the interval  $[-1, 1]$  onto the unit disk  $|\zeta| < 1$  conformally and univalently.

**Lemma 2.** *If a polynomial  $P(z) \in \mathcal{PU}_{n,k}$  then the function*

$$z = F_1(w) := w^{1-n-k} \Phi \left[ \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right) \right]$$

*satisfies the conditions of lemma 1 on the open set*

$$B_1 := \left\{ w : |w| < 1, \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right) \notin [-1, 1] \right\}.$$

**Lemma 3.** *If a polynomial  $P(z) \in \mathcal{PV}_n$  then the function*

$$z = F_2(w) := w^{1-2n-k} \Phi \left[ \frac{1}{2} \left( w^k + \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w^2 + \frac{1}{w^2} \right) \right) \right]$$

*satisfies the conditions of lemma 1 on the open set*

$$B_2 := \left\{ w : |w| < 1, \frac{1}{2} \left( w^k + \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w^2 + \frac{1}{w^2} \right) \right) \notin [-1, 1] \right\}.$$

**Lemma 4.** *If a polynomial  $P(z) \in \mathcal{PW}_n$  then the function*

$$z = F_3(w) := w^{1-2n-k} \Phi \left[ \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w^2 + \frac{1}{w^2} \right) \right) \right]$$

*satisfies the conditions of lemma 1 on the open set*

$$B_3 := \left\{ w : |w| < 1, \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w^2 + \frac{1}{w^2} \right) \right) \notin [-1, 1] \right\}.$$

It follows from the representation of the Chebyshev polynomials (1)-(4) that  $F_1(w) \equiv iw$  for  $P(z) = U_m(T_k(z))$ ,  $F_2(w) \equiv w$  for  $P(z) = V_m(T_k(z))$  and  $F_3(z) \equiv iw$  for  $P(z) = W_m(T_k(z))$ .

### §3. ESTIMATES OF COEFFICIENTS AND COVERING THEOREMS

Every inequality for coefficients of functions from the class  $\mathcal{B}$  implies some inequality for algebraic polynomials.

**Theorem 1.** *Suppose that  $P(z) \in \mathcal{PU}_{n,k}$ . Then the estimate*

$$|c_{n-1}| + |c_n| \leq 2^n$$

*holds. For  $n$  divisible by  $k$  this inequality becomes equality for the polynomial  $U_{n/k}(T_k(z))$ .*

**Proof.** Consider the function  $F_1(w)$  from lemma 2. In some neighborhood of the origin we have

$$F_1(w) = \beta_1 w + \beta_2 w^2 + \beta_3 w^3 + \beta_4 w^4 + \dots$$

It follows from lemma 1 that there exists a function

$$w = f(z) = \alpha_1 z + \alpha_2 z^2 + \alpha_3 z^3 + \dots$$

such that  $f \in \mathcal{B}$  and  $F_1(f(z)) \equiv z$  in the unit disk  $U$ .

We have the following relations

$$\alpha_1 = \frac{1}{\beta_1}, \quad \alpha_2 = -\frac{\beta_2}{\beta_1^3}. \quad (12)$$

On the other hand from the definition of the function  $F_1(w)$  we get

$$\begin{aligned} \beta_1 w^{n+k} + \beta_2 w^{n+k+1} + \beta_3 w^{n+k+2} + \dots &= \Phi \left[ \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right) \right] = \\ &= \Phi \left[ \frac{i}{2} \left( -\frac{c_n}{2^n} \frac{1}{w^{n+k}} - \frac{c_{n-1}}{2^{n-1}} \frac{1}{w^{n+k-1}} + \dots \right) \right] \end{aligned}$$

in some neighborhood of the origin.

Hence

$$\beta_1 = i \frac{2^n}{c_n}, \quad \frac{\beta_2}{\beta_1^3} = i \frac{c_{n-1}}{2^{n-1}}.$$

It follows from (12), (9) that the claimed inequality holds. If  $n$  is divisible by  $k$  and  $P(z) = U_{n/k}(T_k(z))$  then  $F(w) \equiv iw$ . Therefore this inequality becomes equality. The theorem is proved.

The inequality from theorem 1 implies the estimate

$$|c_n| \leq 2^n,$$

which is equivalent to the property of the polynomial  $U_{n/k}(T_k(z))2^{-n}$  to have the least deviation from zero on the interval  $[-1, 1]$  with weight  $\sqrt{1 - T_k^2(z)}$  among polynomials  $P(z) = z^n + \dots$

**Theorem 2.** Suppose that  $P(z) \in \mathcal{PV}_{n,k}$  ( $\mathcal{PW}_{n,k}$ ). Then the estimate

$$|c_n| \leq 2^n$$

holds. For  $n$  divisible by  $k$  this inequality becomes equality for the polynomial  $V_{n/k}(T_k(z))$  (respectively  $W_{n/k}(T_k(z))$ ).

To prove this theorem it is sufficient to consider the functions  $F_2(w)$  and  $F_3(w)$  from lemmas 3 and 4 in place of  $F_1(w)$ .

**Theorem 3.** If  $P(z) \in \mathcal{PU}_{n,k}$  then for any number  $r$ ,  $r > 1$  the image of the ellipse  $|z-1| + |z+1| = r + 1/r$  under the mapping  $\omega = \sqrt{1 - T_k^2(z)}P(z)$  is

a curve lying inside the ellipse with foci  $\pm 1$  and the major axis  $x_{0,r}r^{-n-k+1} + r^{n+k-1}/x_{0,r} \leq r^{-n-k} + r^{n+k}$ , where  $x_{0,r}$  is the root of the equation

$$|c_n|(1-r)^2x = r2^n(1-x)^2, \quad (13)$$

belonging to the interval  $1/r \leq x < 1$ .

For

$$r > r_\lambda = 2\lambda - 1 + 2\sqrt{\lambda(\lambda-1)}, \quad \lambda = 2^n/|c_n|,$$

the image of the ellipse  $|z-1| + |z+1| = r + 1/r$  under the mapping  $\omega = \sqrt{1-T_k^2(z)}P(z)$  is a curve lying outside the ellipse  $|\omega-1| + |\omega+1| = x_{1,r}r^{-n-k+1} + r^{n+k-1}/x_{1,r}$ , where  $x_{1,r}$  is the root of the equation

$$|c_n|(1+r)^2x = r2^n(1+x)^2, \quad (14)$$

belonging to the interval  $1/r \leq x < 1$ .

If  $n$  divisible by  $k$  and  $P(z) = U_{n/k}(T_k(z))$  then  $x_{0,r} = x_{1,r} = 1/r$  and the image of the ellipse  $|z-1| + |z+1| = r + 1/r$  under the mapping  $\omega = \sqrt{1-T_k^2(z)}P(z)$  is the ellipse with foci  $\pm 1$  and the major axis  $r^{-n-k} + r^{n+k}$ .

**Proof.** Consider the function

$$z = F(w) = \beta_1 w + \beta_2 w^2 + \dots$$

from lemma 2. It follows from the proof of theorem 1 that  $\beta_1 = i2^n/c_n$ . Let  $r > 1$  and  $w$  be an arbitrary point on the circle  $|w| = 1/r$ .

If  $|F(w)| \geq 1$  then

$$\left| \Phi \left[ \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right) \right] \right| \geq |w|^{n+k-1} = (1/r)^{n+k-1}. \quad (15)$$

Suppose that  $|F(w)| < 1$ . By lemmas 1 and 2 the point  $w$  is the image of some point  $z \in U$  under the inverse mapping  $w = F_1^{-1}(z) \in \mathcal{B}$ . Using (10) and theorem 1, we get

$$\frac{|z|}{(1-|z|)^2} \geq \frac{2^n}{|c_n|} \frac{1/r}{(1-(1/r))^2} \geq \frac{1/r}{(1-(1/r))^2}. \quad (16)$$



The function  $f(z) = z/(1-z)^2$  is strictly increasing in  $[0, 1)$ . Hence inequalities (16) are equivalent to

$$|z| \geq x_{0,r} \geq 1/r,$$

where  $x_{0,r}$  is the root of the equation (13),  $1/r \leq x_{0,r} < 1$ . Therefore

$$\left| \Phi \left[ \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right) \right] \right| \geq x_{0,r} (1/r)^{n+k-1} \geq (1/r)^{n+k}. \quad (17)$$

Comparing (15) with (17), we get inequality (17) for all  $w$  on the circle  $|w| = 1/r$ . Hence the point  $\frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right)$  lies inside the ellipse with foci  $\pm 1$  and the major axis  $x_{0,r} r^{-n-k+1} + r^{n+k-1}/x_{0,r} \leq r^{-n-k} + r^{n+k}$ . In other words the point  $\sqrt{1 - T_k^2(z)} P(z)$ , where  $z = \frac{1}{2} \left( w + \frac{1}{w} \right)$ , i.e.  $|z - 1| + |z + 1| = r + 1/r$ , lies inside the ellipse with foci  $\pm 1$  and the major axis  $x_{0,r} r^{-n-k+1} + r^{n+k-1}/x_{0,r} \leq r^{-n-k} + r^{n+k}$ .

Now we consider

$$r > r_\lambda = 2\lambda - 1 + 2\sqrt{\lambda(\lambda - 1)}, \quad \lambda = 2^n/|c_n|.$$

Let  $w$ ,  $|w| = 1/r$ ,  $r > r_\lambda$ , be a point of the set  $f(|z| < 1)$ . We have

$$\frac{|F_1(w)|}{(1 + |F_1(w)|)^2} \leq \frac{2^n}{|c_n|} \frac{1/r}{(1 + 1/r)^2} \leq \frac{2^n}{|c_n|} \frac{1/r_\lambda}{(1 + 1/r_\lambda)^2} = \frac{1}{4}.$$

Hence equation (14) has the unique root  $x_{1,r} \in [1/r, 1)$  and this root satisfies the bound

$$|F_1(w)| \leq x_{1,r}. \quad (18)$$

Now we will demonstrate that any point  $w$ ,  $|w| < 1/r_\lambda$ , belongs to the domain  $f(|z| < 1)$ . Suppose the contrary is true. Then

$$r_\lambda < r^* = \inf\{r : r > 1, |F_1(w)| < 1 \ \forall w, |w| = 1/r\}.$$

There is a point  $w^*$  on the circle  $|w| = 1/r^*$  such that

$$|F_1(w^*)| = 1. \quad (19)$$

On the other hand, for any sequence of points  $w_k, |w_k| < 1/r^*, k = 1, 2, \dots$ , converging to  $w^*$  the bound (18) yields

$$|F_1(w_k)| \leq x_{1,r^*}, \quad k = 1, 2, \dots,$$

which contradicts condition (19). It follows from (18) that

$$\left| \Phi \left[ \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right) \right] \right| \leq \frac{x_{1,r}}{r^{n+k-1}}.$$

It is equivalent to the respective statement of the theorem.

The last statement of the theorem is immediately obtained from the representations of Chebyshev polynomials (1), (2). The theorem is proved.

Proofs of the following two theorems are similar.

**Theorem 4.** *If  $P(z) \in \mathcal{PV}_{n,k}$  then for any number  $r$ ,  $r > 1$  the image of the ellipse  $|z - 1| + |z + 1| = r^2 + 1/r^2$  under the mapping  $\omega = \sqrt{(1 + T_k(z))/2} P(z)$  is a curve lying inside the ellipse with foci  $\pm 1$  and the major axis  $x_0 r^{-2n-k+1} + r^{2n+k-1}/x_0 \leq r^{-2n-k} + r^{2n+k}$ , where  $x_{0,r}$  is the root of the equation (13) belonging to the interval  $1/r \leq x < 1$ .*

For

$$r > r_\lambda = 2\lambda - 1 + 2\sqrt{\lambda(\lambda - 1)}, \quad \lambda = 2^n/|c_n|,$$

*the image of the ellipse  $|z - 1| + |z + 1| = r^2 + 1/r^2$  under the mapping  $\omega = \sqrt{(1 + T_k(z))/2} P(z)$ , is a curve lying outside the ellipse  $|\omega - 1| + |\omega + 1| = x_{1,r} r^{-2n-k+1} + r^{2n+k-1}/x_{1,r}$ , where  $x_{1,r}$  is the root of the equation (14) belonging to the interval  $1/r \leq x < 1$ .*

*If  $n$  divisible by  $k$  and  $P(z) = V_{n/k}(T_k(z))$  then  $x_{0,r} = x_{1,r} = 1/r$  and the image of the ellipse  $|z - 1| + |z + 1| = r^2 + 1/r^2$  under the mapping  $\omega = \sqrt{(1 + T_k(z))/2} P(z)$  is the ellipse with foci  $\pm 1$  and the major axis  $r^{-2n-k} + r^{2n+k}$ .*

**Theorem 5.** *If  $P(z) \in \mathcal{PW}_{n,k}$  then for any number  $r$ ,  $r > 1$  the image of the ellipse  $|z - 1| + |z + 1| = r^2 + 1/r^2$  under the mapping  $\omega = \sqrt{(1 - T_k(z))/2} P(z)$  is a curve lying inside the ellipse with foci  $\pm 1$  and the major axis  $x_0 r^{-2n-k+1} + r^{2n+k-1}/x_0 \leq r^{-2n-k} + r^{2n+k}$ , where  $x_{0,r}$  is the root of the equation (13) belonging to the interval  $1/r \leq x < 1$ .*

For

$$r > r_\lambda = 2\lambda - 1 + 2\sqrt{\lambda(\lambda - 1)}, \quad \lambda = 2^n/|c_n|,$$

under the mapping  $\omega = \sqrt{(1 - T_k(z))/2}P(z)$ , the image of the ellipse  $|z-1| + |z+1| = r + 1/r$  is a curve lying outside the ellipse  $|\omega-1| + |\omega+1| = x_{1,r}r^{-2n-k+1} + r^{2n+k-1}/x_{1,r}$ , where  $x_{1,r}$  is a root of the equation (14) from the interval  $1/r \leq x < 1$ .

If  $n$  divisible by  $k$  and  $P(z) = W_{n/k}(T_k(z))$  then  $x_{0,r} = x_{1,r} = 1/r$  and the image of the ellipse  $|z-1| + |z+1| = r^2 + 1/r^2$  under the mapping  $\omega = \sqrt{(1 - T_k(z))/2}P(z)$  is the ellipse with foci  $\pm 1$  and the major axis  $r^{-2n-k} + r^{2n+k}$ .

#### §4. INEQUALITIES FOR DERIVATIVE OF A POLYNOMIAL

**Theorem 6.** If  $P(z) \in \mathcal{PU}_{n,k}$  then

$$\begin{aligned} & k|P(x)T_k(x)T'_k(x) - P'(x)(1 - T_k^2(x))| \leq \\ & \leq (n+k-1 + \sqrt{|c_n/2^n|})|T'_k(x)|\sqrt{1 - (1 - T_k^2(x))P^2(x)}, \quad x \in [-1, 1]. \end{aligned} \quad (20)$$

If  $n$  divisible by  $k$  and  $P(z) = U_{n/k}(T_k(z))$  then this inequality becomes equality for all  $x \in [-1, 1]$ .

**Proof.** Consider the function  $F_1(w)$  from lemma 2 on the set  $B_1$ . The set  $B_1 \setminus \{w : |F_1(w)| = 1\}$  is a family of domains. Denote this family by  $\mathcal{D}_1$ . It follows from lemma 1 that if  $D \in \mathcal{D}_1$  and  $0 \notin D$  then  $F_1(D) \subset \{z : |z| > 1\}$ . If  $0 \in D \in \mathcal{D}_1$ , then  $z = F_1(w)$  maps the domain  $D$  on the unit disk  $U$  conformally and univalently. Suppose that a point  $x \in [-1, 1]$  satisfies the following condition: if  $(w + 1/w)/2 = x$ ,  $|w| = 1$  then  $w$  is a point of regularity of the function  $F_1(w)$ . (All points  $x$  of the interval  $[-1, 1]$  satisfy this condition possibly except for finite number of points). If  $w \in \partial D$ ,  $D \in \mathcal{D}_1$  and  $0 \notin D$ , then at the point  $w$  we have

$$\frac{\partial |F_1|}{\partial |w|} \leq 0. \quad (21)$$

If  $w \in \partial D$ ,  $0 \in D \in \mathcal{D}_1$ , then it follows from (11) that

$$\frac{\partial |F_1|}{\partial |w|} \leq \sqrt{\frac{|c_n|}{2^n}}. \quad (22)$$

Comparing (21) with (22) we get the inequality (22) for all regular points  $w \in \partial D$ ,  $D \in \mathcal{D}_1$ .

If  $|w| = 1$  then the point

$$\omega := \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w + \frac{1}{w} \right) \right)$$

belongs to  $[-1, 1]$ . Hence

$$|\Phi'(\omega)| = \left| 1 \pm \frac{\omega i}{\sqrt{1 - \omega^2}} \right| = \frac{1}{\sqrt{1 - \omega^2}}.$$

A direct computation shows that

$$\begin{aligned} \frac{\partial |F_1|}{\partial |w|} &= 1 - n - k + \frac{1}{\sqrt{1 - \omega^2}} \left| \frac{k}{2} \left( w^k + \frac{1}{w^k} \right) P(x) + \right. \\ &\quad \left. + \frac{1}{4} \left( w^k - \frac{1}{w^k} \right) P'(x) \left( w - \frac{1}{w} \right) \right|. \end{aligned}$$

If  $x = \frac{1}{2} (w + 1/w)$  then from the representation of Chebyshev polynomial we get

$$\frac{1}{2} \left( w^k + \frac{1}{w^k} \right) = T_k(x), \quad \frac{i}{2} \left( w^k - \frac{1}{w^k} \right) = \sqrt{1 - T_k^2(x)}, \quad T'_k(x) = \frac{k \left( w^k - \frac{1}{w^k} \right)}{\left( w - \frac{1}{w} \right)}.$$

Using the last equalities we obtain the inequality (20).

If  $n$  is divisible by  $k$  and  $P(z) = U_{n/k}(T_k(z))$  than  $F(w) \equiv iw$ . Therefore this inequality becomes the equality. The theorem is proved.

**Theorem 7.** *If  $P(z) \in \mathcal{PV}_{n,k}$  then*

$$\begin{aligned} &\sqrt{1 - x^2} |2P'(x)(1 + T_k(x)) + T'_k(x)P(x)| \leq \\ &\leq (2n + k - 1 + \sqrt{|c_n/2^n|}) \sqrt{(1 + T_k(x))(2 - (1 + T_k(x))P^2(x))}, \quad x \in [-1, 1]. \end{aligned}$$

*If  $n$  divisible by  $k$  and  $P(z) = V_{n/k}(T_k(z))$  then this inequality becomes equality for all  $x \in [-1, 1]$ .*

**Proof.** Consider the function  $F_2(w)$  from lemma 3 on the set  $B_2$ . Let the set  $B_2 \setminus \{w : |F_2(w)| = 1\}$  is a family of domains. Denote this family by  $\mathcal{D}_2$ . It follows from lemma 1 that if  $D \in \mathcal{D}_2$  and  $0 \notin D$  then  $F_2(D) \subset \{z : |z| > 1\}$ .

If  $0 \in D \in \mathcal{D}_2$ , then  $z = F_2(w)$  maps the domain  $D$  on the unit disk  $U$  conformally and univalently. Suppose that a point  $x \in [-1, 1]$  satisfies the following condition: if  $(w^2 + 1/w^2)/2 = x$ ,  $|w| = 1$  then  $w$  is a point of regularity of the function  $F_2(w)$ . (All points  $x$  of the interval  $[-1, 1]$  satisfy this condition possibly except for finite number of points). If  $w \in \partial D$ ,  $D \in \mathcal{D}_2$  and  $0 \notin D$ , then at the point  $w$  we have

$$\frac{\partial |F_2|}{\partial |w|} \leq 0. \quad (23)$$

If  $w \in \partial D$ ,  $0 \in D \in \mathcal{D}_2$ , then it follows from (11) that

$$\frac{\partial |F_2|}{\partial |w|} \leq \sqrt{\frac{|c_n|}{2^n}}. \quad (24)$$

Comparing (23) with (24) we get the inequality (24) for all regular points  $w \in \partial D$ ,  $D \in \mathcal{D}_2$ .

If  $|w| = 1$  then the point

$$\omega := \frac{1}{2} \left( w^k + \frac{1}{w^k} \right) P \left( \frac{1}{2} \left( w^2 + \frac{1}{w^2} \right) \right)$$

belongs to  $[-1, 1]$ . Hence

$$|\Phi'(\omega)| = \left| 1 \pm \frac{\omega i}{\sqrt{1 - \omega^2}} \right| = \frac{1}{\sqrt{1 - \omega^2}}.$$

A direct computation shows that

$$\begin{aligned} \frac{\partial |F_2|}{\partial |w|} &= 1 - 2n - k + \frac{1}{\sqrt{1 - \omega^2}} \left| \frac{k}{2} \left( w^k - \frac{1}{w^k} \right) P(x) + \right. \\ &\quad \left. + \frac{1}{2} \left( w^k + \frac{1}{w^k} \right) P'(x) \left( w^2 - \frac{1}{w^2} \right) \right|. \end{aligned}$$

We have

$$x = \frac{1}{2} \left( w^2 + \frac{1}{w^2} \right).$$

Therefore

$$T_k(x) = \frac{1}{2} \left( w^{2k} + \frac{1}{w^{2k}} \right) = \frac{1}{2} \left( w^k \pm \frac{1}{w^k} \right)^2 \mp 1.$$

Hence

$$\left( w^k \pm \frac{1}{w^k} \right) = \sqrt{2} \sqrt{T_k(x) \pm 1}, \quad T'_k(x) = \frac{k \left( w^{2k} - \frac{1}{w^{2k}} \right)}{\left( w^2 - \frac{1}{w^2} \right)}.$$

Using the last equalities we get inequality from the statement of the theorem.

If  $n$  is divisible by  $k$  and  $P(z) = V_{n/k}(T_k(z))$  than  $F(w) \equiv w$ . Therefore these inequalities become equalities. The theorem is proved.

**Theorem 8.** *If  $P(z) \in \mathcal{PW}_{n,k}$  then*

$$\begin{aligned} & \sqrt{1-x^2}|2P'(x)(T_k(x)-1) + T'_k(x)P(x)| \leq \\ & \leq (2n+k-1 + \sqrt{|c_n/2^n|})\sqrt{(1-T_k(x))(2+(T_k(x)-1)P^2(x))}, \quad x \in [-1, 1]. \end{aligned}$$

*For  $n$  divisible by  $k$  this inequality becomes equality for the polynomial  $W_{n/k}(T_k(z))$*

**Remark 1.** If we assume  $k = 1$  in our statements then we obtain some results from articles [7], [8].

**Remark 2.** If we assume  $k = 2$  in the theorems 2, 4, 7 then we get statements for polynomials with restriction  $|P(x)||x| \leq 1$  on the interval  $[-1, 1]$ .

#### ACKNOWLEDGMENTS

This research was carried out with the financial support of the Russian Foundation for Basic Research (grant no.08-01-00028), the Programme of Support of Leading Scientific Schools of RF (grant no. 2810.2008.1) and the Far-Eastern Branch of RAS (grant no. 09-I-P4-02).

#### References

- [1] P. Borwein, T. Erdélyi, *Polynomials and polynomial inequalities*, Grad. Texts in Math., 161, Springer-Verlag, New York, 1995.
- [2] G. Min, Inequalities for rational functions with prescribed poles, *Can. J. Math.*, 50, No 1, 152-166 (1998).
- [3] Q.I. Rahman, On a problem of Turan about polynomials with curved majorants, *Trans. Amer. Math. Soc.*, 163, 447-455 (1972).

- [4] Q.I. Rahman, G. Schmeisser, Markoff type inequalities for curved majorants, *Numerical methods of approximation theory*, 8, 169-183 (1987).
- [5] M.A. Lachance, Bernstein and Markov inequalities for constrained polynomials, *Lect. Notes Math.*, 1045, 125-135 (1984).
- [6] V.N. Dubinin, Conformal mappings and inequalities for polynomials, *Algebra and analysis*, 13, No 5, 16-43 (2001).
- [7] V.N. Dubinin, A.V. Olesov, Application Of Conformal Mappings To Inequalities For Polynomials, *Journal of Mathematical Sciences*, 122, No 6, 3630-3640 (2004).
- [8] V.N. Dubinin, S.I. Kalmykov, Extremal properties of Chebyshev polynomials, *Far Eastern Mathematical Journal*, 5, No 2, 169-177 (2004).
- [9] V.I. Lebedev, *Functional analysis and calculus mathematics*, Nauka, Moscow, 2000.
- [10] N.A. Lebedev, *Method of areas in univalent function theory*, Nauka, Moscow, 1975.
- [11] G.M. Goluzin, *Geometric theory of functions of complex variable*, Nauka, Moscow 1966; English transl., Transl. Math. Monogr., vol. 26, Amer. Math. Soc., Providence, RI, 1969.

# POLYGONAL APPROXIMATIONS OF NON-RECTIFIABLE CURVES AND THE JUMP PROBLEM

B.A. KATS

Department of Mathematics, Kazan State University, Kremlevskaya Street, 18,  
Kazan, Tatarstan, Russia 420008

Chair of Mathematics, Kazan State Architecture and Civil Engineering Univer-  
sity, Zelenaya Street, 1, Kazan, Tatarstan, Russia 420043

E- mail address: architec@mi.ru

**ABSTRACT.** We consider so called jump problem, i.e. the problem on evaluation of holomorphic in  $\mathbb{C} \setminus \Gamma$  function with given difference of its boundary values on curve  $\Gamma$ . If this curve is rectifiable, then solution of the problem is representable as the Cauchy integral over  $\Gamma$ . The present paper is dealing with the case of non-rectifiable path  $\Gamma$ . Generally speaking, in this situation the integral over  $\Gamma$  does not exist, but the problem can be solved in terms of polygonal approximation of this curve.

**Key words and phrases:** holomorphic function, jump problem, non-rectifiable curve, approximation, fractal.

## 1. INTRODUCTION

We consider the following boundary value problem for holomorphic functions. Let  $\Gamma$  be a closed Jordan curve on the complex plane  $\mathbb{C}$  bounding finite domain  $D^+$ , and  $D^- = \mathbb{C} \setminus D^+$ . A function  $f(t)$  is defined on  $\Gamma$ . We seek a holomorphic in  $\mathbb{C} \setminus \Gamma$  function  $\Phi(z)$  such that  $\Phi(\infty) = 0$ , the boundary values  $\lim_{D^+ \ni z \rightarrow t} \Phi(z) \equiv \Phi^+(t)$  and  $\lim_{D^- \ni z \rightarrow t} \Phi(z) \equiv \Phi^-(t)$  exist for any  $t \in \Gamma$ , and

$$\Phi^+(t) - \Phi^-(t) = f(t), t \in \Gamma. \quad (1)$$

The problem is called jump problem. It is well known and has numerous applications in elasticity theory and so on (see, for instance, [12, 4]). If the curve  $\Gamma$  is piecewise-smooth and the jump  $f(t)$  satisfies the Hölder condition

$$\sup \left\{ \frac{|f(t') - f(t'')|}{|t' - t''|^\nu} : t', t'' \in \Gamma, t' \neq t'' \right\} \equiv h_\nu(f, \Gamma) < \infty \quad (2)$$

with exponent  $\nu \in (0, 1]$ , then unique solution of this problem is the Cauchy integral

$$\Phi(z) = \frac{1}{2\pi i} \int_{\Gamma} \frac{f(\zeta) d\zeta}{\zeta - z}. \quad (3)$$

Below we denote  $H_\nu(\Gamma)$  the set of all functions satisfying (2).

If the curve  $\Gamma$  is rectifiable but non-smooth, then the Cauchy integral (3) has boundary values  $\Phi^\pm$  if  $f$  satisfies the Hölder condition with exponent  $\nu > \frac{1}{2}$  (see [2, 13]). It is not difficult to show that this boundary values satisfy the equality (1), i.e. the jump problem on non-smooth rectifiable curve is solvable if the Hölder



B.A. KATS

exponent of the jump exceeds  $\frac{1}{2}$ . As shown in [2], this result is cannot be improved in the whole class of rectifiable curves.

Now let  $\Gamma$  be non-rectifiable. Then the jump problem (1) keeps the sense, but the Cauchy integral over that curve loses definiteness.

The author researched the jump problem for non-rectifiable curves (see, for instance, [5, 6, 7]) and obtained certain conditions of its solvability in terms of various fractal dimensions of curve  $\Gamma$ . In particular, he proved [6] that the jump problem on non-rectifiable curve is solvable under assumption

$$\nu > \frac{\text{Dm } \Gamma}{2}. \quad (4)$$

Here  $\text{Dm } \Gamma$  means upper metric dimension of the set  $\Gamma$ , i.e.

$$\text{Dm } \Gamma = \limsup_{\varepsilon \rightarrow 0} \frac{\log N(\varepsilon)}{-\log \varepsilon},$$

where  $N(\varepsilon)$  stands for the least number of disks of diameter  $\varepsilon$  necessary for covering of the set  $\Gamma$  (see [10, 3]). The bound (4) cannot be improved on the class of curves of fixed upper metric dimension. Note that in general the solution of problem (1) on non-rectifiable curve cannot be represented by the Cauchy integral.

In the present paper we study the jump problem by means of polygonal approximations of non-rectifiable curve  $\Gamma$ . The scheme of that approach is rather easy. Let  $\Gamma_1, \Gamma_2, \dots, \Gamma_n, \dots$  be a sequence of polygons approximating  $\Gamma$  in some sense. We extend the jump  $f$  from  $\Gamma$  onto the whole complex plane and consider the limit of Cauchy integrals

$$\Phi_*(z) = \lim_{n \rightarrow \infty} \frac{1}{2\pi i} \int_{\Gamma_n} \frac{f^\varepsilon(\zeta) d\zeta}{\zeta - z}, \quad (5)$$

where  $f^\varepsilon$  is extension of the jump  $f$ . The integrals are defined here because the polygons  $\Gamma_n$  are piecewise-smooth. If the limit exists and has boundary values  $\Phi_*^\pm$  satisfying (1), then it is a solution of the jump problem. This approach enables us to obtain an approximate solution of the jump problem with a bound for its error.

A non-rectifiable curve can be approximated by polygons in various ways. In the next section we are dealing with so called monotone approximations. Then we consider representability of the solution (5) by the Cauchy integral (Section 3) and certain examples and commentaries (Section 4).

## 2. MONOTONE POLYGONAL APPROXIMATIONS

We use the following notation. The width  $w(\delta)$  of finite domain  $\delta$  is diameter of the most open disk lying in  $\delta$ . The Hausdorff distance between sets  $A$  and  $B$  is defined by equality

$$\text{dist}_H(A, B) := \sup\{\text{dist}(z, B) : z \in A\} + \sup\{\text{dist}(z, A) : z \in B\}.$$

If  $\gamma$  is rectifiable curve, then  $|\gamma|$  stands for its length.

**Definition 1.** We call a sequence of closed polygonal lines  $\Gamma_1, \Gamma_2, \dots, \Gamma_n, \dots$  increasing (or decreasing) approximation of  $\Gamma$  if these lines bound finite domains  $D_1^+, D_2^+, \dots, D_n^+, \dots$  such that  $D_n^+ \subset D_{n+1}^+ \subset D^+$  (correspondingly,  $D_n^+ \supset D_{n+1}^+ \supset D^+$ ) for any positive integer  $n$  and  $\lim_{n \rightarrow \infty} \text{dist}_H(\Gamma_n, \Gamma) = 0$ .

We call a sequence  $\Gamma_1, \Gamma_2, \dots, \Gamma_n, \dots$  monotone approximation of  $\Gamma$  if it is either increasing or decreasing approximation of this curve. If a sequence of polygons  $\{\Gamma_n\}$  approximates a non-rectifiable curve, then  $|\Gamma_n| \rightarrow \infty$ .

Let a sequence  $\{\Gamma_n\}$  be monotone polygonal approximation of  $\Gamma$ . We put  $\Delta_n = D_{n+1}^+ \setminus \overline{D_n^+}$  if this approximation increases, and  $\Delta_n = D_n^+ \setminus \overline{D_{n+1}^+}$  if it decreases,  $n = 1, 2, \dots$ . As  $D_n^+$  and  $D_{n+1}^+$  are polygonal domains, since  $\Delta_n$  is either polygonal domain (maybe, double connected) or union of finite number of disjoint polygonal domains. Let  $\lambda_n$  stand for sum of perimeters of all connected components of  $\Delta_n$ , and  $\omega_n$  for maximal width of these components.

In what follows we use also the Whitney extension operator (see, for instance, [14]). If a function  $f(t)$  is defined on a compact set  $K \subset \mathbb{C}$  and satisfies the Hölder condition with exponent  $\nu$ , i.e.  $f \in H_\nu(K)$ , then its Whitney extension  $f^w(z)$  is defined on the whole complex plane  $\mathbb{C}$  and satisfies the Hölder condition with the same exponent  $\nu$ , i.e.  $f^w \in H_\nu(\mathbb{C})$ . Moreover,  $h_\nu(f^w, \mathbb{C}) = h_\nu(f, K)$ . In addition,  $f^w(x + iy)$  has partial derivatives of all orders at any point  $x + iy = z \in \mathbb{C} \setminus K$  and  $|\nabla f^w(z)| \leq h_\nu(f, K) \text{dist}^{\nu-1}(z, \Gamma)$ .

**Lemma 1.** *Let  $\delta$  be finite domain with Jordan rectifiable boundary  $\gamma$ ,  $f \in H_\nu(\gamma)$ , and  $f^w$  is the Whitney extension of the function  $f$  from the curve  $\gamma$ . If  $p < \frac{1}{1-\nu}$ , then*

$$\iint_{\delta} |\nabla f^w|^p dx dy \leq C h_\nu^p(f, \gamma) |\gamma| w^{1-p(1-\nu)}(\delta),$$

where  $C \leq 4\pi(1 - 2^{p(1-\nu)-1})^{-1}$ .

*Proof.* We consider the Whitney decomposition of domain  $\delta$  (see [14]). It consists of dyadic squares  $Q$  such that  $a(Q) \leq \text{dist}(z, \gamma) \leq 4a(Q)$  for any  $z \in Q$ , where  $a(Q)$  is length of side of  $Q$ . Let  $m_n$  stand for number of squares  $Q$  such that  $a(Q) = 2^{-n}$ . Then the cited above bound for  $|\nabla f^w(z)|$  yields inequality

$$\iint_{\delta} |\nabla f^w|^p dx dy \leq h_\nu^p(f, \gamma) \sum_{n=-\infty}^{\infty} 2^{-n(2-p(1-\nu))} m_n.$$

All  $m_n$  squares with side  $2^{-n}$  are situated in  $4 \cdot 2^{-n}$ -neighborhood of the curve  $\gamma$ . The square of the neighborhood does not exceed  $4\pi|\gamma|2^{-n}$ . Hence,  $m_n \leq 4\pi|\gamma|2^n$ . But  $m_n = 0$  for  $w(\delta) < 2^{-n}$ . Whence,

$$\iint_{\delta} |\nabla f^w|^p dx dy \leq 4\pi h_\nu^p(f, \gamma) |\gamma| \sum_{2^{-n} \leq w(\delta)} 2^{-n(1-p(1-\nu))}.$$

The series converges under assumptions of the lemma and

$$\sum_{2^{-n} \leq w(\delta)} 2^{-n(1-p(1-\nu))} \leq w^{1-p(1-\nu)}(\delta) \sum_{n=0}^{\infty} 2^{n(p(1-\nu)-1)} = \frac{w^{1-p(1-\nu)}(\delta)}{1 - 2^{p(1-\nu)-1}},$$

what concludes the proof.

**Theorem 1.** *If non-rectifiable curve  $\Gamma$  has a monotone polygonal approximation  $\{\Gamma_n\}$  such that*

$$\sum_{n=1}^{\infty} \lambda_n \omega_n^{1-p(1-\nu)} < \infty \quad (6)$$

B.A. KATS

for certain  $\nu \in (0, 1)$  and  $p > 2$ , then any function  $f \in H_\nu(\Gamma)$  has an extension  $f^\mathcal{E}$  such that equality (5) defines a solution of the jump problem (1).

*Proof.* We extend the jump  $f$  in the following way. First we apply the Whitney extension operator and obtain an extended function  $f^w(z)$ . Then we put  $\Omega := \cap_{n \geq 1} \Gamma_n$ , restrict  $f^w$  on compact  $\bar{\Omega}$  and apply again the Whitney operator to this restriction. As a result of this double Whitney extension we obtain a function  $f^\mathcal{E}$ . It has all mentioned above properties of the function  $f^w$ , but additionally it has the following property:

– restriction of the function  $f^\mathcal{E}$  on any connected component  $\delta$  of  $\mathbb{C} \setminus \bar{\Omega}$  equals there to the Whitney extension of the restriction of  $f^w$  on the boundary  $\partial\delta$  of this component.

This property follows from the construction of the Whitney extension operator (see [14]). Below we call it *restrictive property*.

Let us consider an increasing polygonal approximation  $\{\Gamma_n\}$ . Obviously,

$$\Phi_n(z) := \frac{1}{2\pi i} \int_{\Gamma_n} \frac{f^\mathcal{E}(\zeta) d\zeta}{\zeta - z} = \sum_{j=0}^n \frac{1}{2\pi i} \int_{\partial\Delta_j} \frac{f^\mathcal{E}(\zeta) d\zeta}{\zeta - z},$$

where  $\Delta_0 \equiv D_1^+$ . The first partial derivatives of  $f^\mathcal{E}$  are integrable in any connected component of  $\Delta_j, j = 0, 1, \dots, n$ , by virtue of the restrictive property and Lemma 1. Therefore we can apply the Cauchy–Green formula to each term of the last sum and obtain

$$\Phi_n(z) = \sum_{j=0}^n \left( \chi(\Delta_j, z) f^\mathcal{E}(z) + \frac{1}{2\pi i} \iint_{\Delta_j} \frac{\partial f^\mathcal{E}(\zeta)}{\partial \bar{\zeta}} \frac{d\zeta d\bar{\zeta}}{\zeta - z} \right).$$

Here and in what follows  $\chi(A, z)$  means characteristic function of set  $A$ , i.e.  $\chi(A, z) = 1$  for  $z \in A$  and  $\chi(A, z) = 0$  for  $z \in \mathbb{C} \setminus A$ . Hence,

$$\Phi_n(z) = \chi(D_n^+, z) f^\mathcal{E}(z) + \frac{1}{2\pi i} \iint_{D_n^+} \frac{\partial f^\mathcal{E}(\zeta)}{\partial \bar{\zeta}} \frac{d\zeta d\bar{\zeta}}{\zeta - z}.$$

According to the restrictive property and Lemma 1 the derivative  $\frac{\partial f^\mathcal{E}(\zeta)}{\partial \bar{\zeta}}$  is integrable in  $D^+$ . Consequently, the limit (5) exists and

$$\Phi_*(z) = \chi(D^+, z) f^\mathcal{E}(z) + \frac{1}{2\pi i} \iint_{D^+} \frac{\partial f^\mathcal{E}(\zeta)}{\partial \bar{\zeta}} \frac{d\zeta d\bar{\zeta}}{\zeta - z}. \quad (7)$$

The term  $\chi(D^+, z) f^\mathcal{E}(z)$  has jump  $f$  on  $\Gamma$ . The derivative  $\frac{\partial f^\mathcal{E}(\zeta)}{\partial \bar{\zeta}}$  under assumption (6) is integrable in  $D^+$  with degree  $p > 2$ . Consequently, the integral term of (7) is continuous in the whole complex plane (see, for instance, [?]). Thus, the theorem is proved for increasing approximations. The proof for decreasing approximations is the same.

Obviously,  $\lambda_n \leq |\Gamma_n| + |\Gamma_{n+1}|$  and  $\omega_n \leq \text{dist}_H(\Gamma_n, \Gamma)$ . Hence, there is valid

**Corollary 1.** *If non-rectifiable curve  $\Gamma$  has a monotone polygonal approximation  $\{\Gamma_n\}$  such that*

$$\sum_{n=1}^{\infty} (|\Gamma_n| + |\Gamma_{n+1}|) \text{dist}_H^{1-p(1-\nu)}(\Gamma_n, \Gamma) < \infty \quad (8)$$

for certain  $\nu \in (0, 1)$  and  $p > 2$ , then any function  $f \in H_\nu(\Gamma)$  has an extension  $f^\mathcal{E}$  such that equality (5) defines a solution of the jump problem (1).

Moreover, we can change the sequence  $\{\Gamma_n\}$  by its subsequence with prescribed decrease of  $\text{dist}_H(\Gamma_n, \Gamma)$ , for instance,  $\text{dist}_H(\Gamma_n, \Gamma) \leq n^{-\alpha}$ . If  $|\Gamma_n| \leq cn^\beta$ , then the series (8) converges for  $\alpha(1 - p(1 - \nu)) - \beta > 1$ , and condition  $p > 2$  implies  $\nu > \frac{\alpha + \beta + 1}{2\alpha}$ . The left side is less than one if  $\alpha > \beta + 1$ . Thus, we obtain

**Corollary 2.** *Let non-rectifiable curve  $\Gamma$  have a monotone polygonal approximation  $\{\Gamma_n\}$  such that  $\text{dist}_H(\Gamma_n, \Gamma) \leq n^{-\alpha}$  and  $|\Gamma_n| \leq cn^\beta$  for  $n = 1, 2, \dots$ . If  $\alpha > \beta + 1$  and*

$$\nu > \frac{\alpha + \beta + 1}{2\alpha},$$

*then any function  $f \in H_\nu(\Gamma)$  has an extension  $f^\mathcal{E}$  such that equality (5) defines a solution of the jump problem (1).*

The following corollary can be proved in just the same way.

**Corollary 3.** *Let non-rectifiable curve  $\Gamma$  have a monotone polygonal approximation  $\{\Gamma_n\}$  such that  $\text{dist}_H(\Gamma_n, \Gamma) \leq 2^{-n}$  and  $|\Gamma_n| \leq c \cdot 2^{\alpha n}$  for  $n = 1, 2, \dots$ . If  $\alpha < 1$  and*

$$\nu > \frac{\alpha + 1}{2},$$

*then any function  $f \in H_\nu(\Gamma)$  has an extension  $f^\mathcal{E}$  such that equality (5) defines a solution of the jump problem (1).*

Now let us consider  $\Phi_n$  as approximation of the solution  $\Phi_*$  and bound the error, i.e. the difference  $\Phi_* - \Phi_n$ . Obviously, this difference cannot be small in domain  $R_n^+$  equaling to  $D^+ \setminus \overline{D_n^+}$  for increasing approximation and to  $D_n^+ \setminus \overline{D^+}$  for decreasing one. Therefore, the intrinsic definition of the error is

$$E_n := \sup\{|\Phi_*(z) - \Phi_n(z)| : z \in R_n^-\},$$

where  $R_n^- := \mathbb{C} \setminus \overline{R_n^+}$ . According to (7), under assumption of Theorem 1 we have

$$\Phi_*(z) - \Phi_n(z) = \pm \frac{1}{2\pi i} \iint_{R_n^+} \frac{\partial f^\mathcal{E}(\zeta)}{\partial \bar{\zeta}} \frac{d\zeta d\bar{\zeta}}{\zeta - z}, z \in R_n^-,$$

where the sign is plus for increasing approximation and minus for decreasing one. Hence,

$$E_n \leq \frac{1}{\pi} \left( \iint_{R_n^+} \left| \frac{\partial f^\mathcal{E}}{\partial \bar{\zeta}}(\zeta) \right|^p dxdy \right)^{1/p} \left( \iint_{R_n^+} |\zeta - z|^{-q} dxdy \right)^{1/q},$$

where  $p^{-1} + q^{-1} = 1$ ,  $z = x + iy$  and  $z \in R_n^-$ . According to Lemma 1

$$\iint_{R_n^+} \left| \frac{\partial f^\mathcal{E}}{\partial \bar{\zeta}}(\zeta) \right|^p dxdy \leq Ch_\nu^p(f, \Gamma) \sum_{j=n+1}^{\infty} \lambda_n \omega_n^{1-p(1-\nu)},$$

where  $C$  is constant from the lemma. If  $d$ ,  $d_n$  and  $r_n$  are diameters of  $\Gamma$ ,  $\Gamma_n$  and  $R_n^+$  correspondingly, then  $r_n = d$  for increasing approximation and  $r_n = d_n$  for decreasing one. As  $q < 2$ , since

$$\iint_{R_n^+} |\zeta - z|^{-q} dxdy \leq \iint_{|\zeta - z| \leq r_n} |\zeta - z|^{-q} dxdy \leq \frac{2\pi}{2 - q} r_n^{2-q}.$$

If the approximation decreases, then  $d_n \leq 2d$  for sufficiently large  $n$ . Thus, there is valid

**Theorem 2.** *If all assumptions of Theorem 1 are fulfilled, then*

$$E_n \leq M h_\nu(f, \Gamma) d^{(p-2)/p} \left( \sum_{j=n+1}^{\infty} \lambda_n \omega_n^{1-p(1-\nu)} \right)^{1/p},$$

where  $M \leq 2^{(q+1)/q} (1 - 2^{p(1-\nu)-1})^{-1/p} (2 - q)^{-1/q}$ . If the approximation increases, then this bound fulfils for any  $n$ , and for decreasing approximation it is valid for all  $n$  such that  $d_n \leq 2d$ .

This theorem means that under assumptions of Corollary 2 we have  $E_n = O(n^{-\mu+\varepsilon})$ , where  $\mu = 2\nu\alpha - (\alpha + \beta + 1)$ , and under assumptions of Corollary 3  $E_n = O(2^{-(\kappa-\varepsilon)n})$ , where  $\kappa = 2\nu - (\alpha + 1)$ . Here  $\varepsilon$  stands for arbitrarily small positive value.

### 3. INSCRIBED POLYGONS AND CAUCHY INTEGRAL

Generally speaking, the limit (5) cannot be represented by the Cauchy integral because the integral  $\int_\Gamma f dz$  over non-rectifiable path  $\Gamma$  is not defined. It is of interest, that the representability of  $\Phi_*$  by the Cauchy integral is related with polygons which are inscribed into the curve  $\Gamma$ .

We consider all finite sequences  $\tau = \{t_1, t_2, \dots, t_n\}$  of points on the curve  $\Gamma$  numbered in order of path-tracing. That sequence determines polygon  $P_\tau$  consisting of segments  $[t_j, t_{j+1}]$ ,  $j = 1, 2, \dots, n$  (here and below  $t_{n+1} \equiv t_1$ ). This polygon is inscribed into  $\Gamma$ ; it may be self-intersecting. Let  $\Psi(x)$  be real continuous increasing function defined for  $x \geq 0$  so that  $\Psi(0) = 0$  and  $\lim_{x \rightarrow \infty} \Psi(x) = \infty$ . By  $\psi(x)$  we denote its inverse function. We relate with an inscribed polygon  $P_\tau$  the following sum:

$$\sigma_\Psi(P_\tau) := \sum_{j=1}^n \Psi(|t_{j+1} - t_j|).$$

**Definition 2.** *A path  $\Gamma$  is called  $\Psi$ -rectifiable if there exists a constant  $C > 0$  such that  $\sigma_\Psi(P_\tau) \leq C$  for any inscribed polygon  $P_\tau$ .*

If  $\Psi(x) \equiv x$ , then any  $\Psi$ -rectifiable curve is rectifiable in usual sense. But if, for instance,  $\Psi(x) = x^p$ ,  $p > 1$ , then  $\Psi$ -rectifiable curve can be non-rectifiable.

Let  $z = \xi(x)$  be a mapping of segment  $I = [0, 1]$  onto the curve  $\Gamma$ . Then  $\sigma_\Psi(P_\tau) := \sum_{j=1}^n \Psi(|\xi(x_{j+1}) - \xi(x_j)|)$ , where  $\{x_j\}$  is increasing sequence of points on  $I$ . A function  $\xi(x)$  is called function with bounded  $\psi$ -variation if the sums  $\sum_{j=1}^n \Psi(|\xi(x_{j+1}) - \xi(x_j)|)$  are bounded (see [15, 11]). Hence,  $\Gamma$  is  $\Psi$ -rectifiable if and only if the mapping  $\xi$  has bounded  $\Psi$ -variation. This fact enables us to apply L.C. Young's theory of Stieltjes integral (see [15, 11]). As shown in [8], the Cauchy integral in the Stieltjes form

$$\frac{1}{2\pi i} \int_\Gamma f(\zeta) d \log(\zeta - z)$$

exists by virtue of the Young theorem [15] if  $f$  has bounded  $\Theta$ -variation and  $\sum_{n=1}^{\infty} \psi(\frac{1}{n}) \theta(\frac{1}{n}) < \infty$ ; here  $\Theta$  is a function satisfying all our assumption on the function  $\Psi$ , and  $\theta$  is its inverse function. In particular, if  $\Gamma$  is  $\Psi$ -rectifiable and

$f \in H_\nu(\Gamma)$ , then  $f$  has bounded  $\Theta$ -variation for  $\Theta(x) = \Psi((\frac{x}{h})^{1/\nu})$ ,  $h = h_\nu(f, \Gamma)$ . In this case  $\theta(x) = h\psi^\nu(x)$ , and the Cauchy integral exists in the Stieltjes sense if

$$\sum_{n=1}^{\infty} \psi^{1+\nu} \left( \frac{1}{n} \right) < \infty. \quad (9)$$

Moreover, there is valid the following result.

**Lemma 2.7**[8] *Let  $\Gamma$  be  $\Psi$ -rectifiable curve, where the function  $\Psi$  is convex and satisfies the condition (9). If a function  $f \in H_\nu(\Gamma)$  has an extension  $F \in H_\nu(\mathbb{C})$  and  $F$  has first partial derivatives which are locally integrable on the complex plane, then there is valid the Cauchy–Green formula*

$$\frac{1}{2\pi i} \int_{\Gamma} f(\zeta) d \log(\zeta - z) = \chi(D^+, z) F(z) + \frac{1}{2\pi i} \iint_{D^+} \frac{\partial F(\zeta)}{\partial \bar{\zeta}} \frac{d\zeta d\bar{\zeta}}{\zeta - z}.$$

Our double Whitney extension  $f^\varepsilon$  satisfies assumptions of this lemma. Thus, we obtain from the last lemma and equality (7)

**Theorem 3.** *Let  $\Gamma$  be a  $\Psi$ -rectifiable curve, where the function  $\Psi$  is convex and satisfies the condition (9). If it has a monotone polygonal approximation satisfying condition (6), then the jump problem (1) has a solution (5) for any  $f \in H_\nu(\Gamma)$ , and this solution is representable as the Cauchy integral*

$$\Phi_*(z) = \frac{1}{2\pi i} \int_{\Gamma} f(\zeta) d \log(\zeta - z),$$

where integration is understood in the Stieltjes sense.

#### 4. EXAMPLES AND COMMENTARY

4.1. We consider first so called von Koch snowflake. This well known self-similar fractal curve bounds domain  $D^+ = T_0 \cup (\bigcup_{n=1}^{\infty} \bigcup_{j=1}^{3 \cdot 4^{n-1}} T_{nj})$ , where  $T_0$  is regular triangle with unit sides and  $T_{nj}$  are similar regular triangles with sides  $1/3^n$ . We put  $D_N^+ := T_0 \cup (\bigcup_{n=1}^N \bigcup_{j=1}^{3 \cdot 4^{n-1}} T_{nj})$ ,  $\Gamma_N := \partial D_N^+$ . Obviously, pre-fractals  $\Gamma_N$  forms increasing approximation of  $\Gamma$ . The difference  $\Delta_N = D_{N+1}^+ \setminus D_N^+$  consists of  $3 \cdot 4^N$  triangles with side  $3^{-N-1}$ . Hence, its perimeter is  $\lambda_N = (4/3)^N$ , and its width is  $\omega_N = c \cdot 3^{-N}$ ,  $c = \sqrt{3}/9$ . The series (6) converges for  $\nu > \log_3 2$ . The fractal dimension of the von Koch snowflake equals to  $\log_4 2 = 2 \log_3 2$ . Thus, for this curve the condition (6) for solvability of the jump problem coincides with (4).

4.2. The following example shows that the condition (6) can improve (4). Let  $\beta > 1$ . We put  $K_n = 2^{[n\beta]}$ , where square brackets mean entire part, and divide the segment  $[2^{-n}, 2^{-n+1}]$  of real axis into  $K_n$  equal parts of length  $\alpha_n = 2^{-n}/K_n$  each one. We denote by  $x_{n,j}$  the ends of these parts, i. e.  $x_{n,j} = 2^{-n} + j\alpha_n$ ,  $j = 0, 1, \dots, 2^{[n\beta]} - 1$ , and consider vertical segments  $I_{n,j} = [x_{n,j}, x_{n,j} + i2^{-n}]$ . Then we fix a decreasing sequence of positive numbers  $\varepsilon_n$  such that  $\varepsilon_n < \frac{1}{2}\alpha_n$ ,  $n = 1, 2, \dots$ , and consider mutually disjoint rectangles  $\delta_{n,j} = \{z = x + iy : x_{n,j} < x < x_{n,j} + \varepsilon_n, 0 < y < 2^{-n}\}$ . Let  $\delta_0$  be square  $\{z = x + iy : 0 < x < 1, 0 < y < 1\}$  and  $D^+ \equiv \delta_0 \setminus (\bigcup_{n=1}^{\infty} \bigcup_{j=0}^{2^{[n\beta]}-1} \delta_{n,j})$ , i. e.  $D^+$  is unit square with countable set of rectangular cuts condensing to origin. We denote by  $\Gamma^*$  the boundary of domain

B.A. KATS

$D^+$ . It is non-rectifiable. As shown in [5, 6],

$$\text{Dm } \Gamma^* = \frac{2\beta}{\beta + 1} \quad (10)$$

in the case  $\varepsilon_n = \frac{1}{2}\alpha_n$ . But the considerations of these papers keep correctness for  $\varepsilon_n < \frac{1}{2}\alpha_n$ , too. Thus, the equality (10) is valid under assumptions of the present example.

The polygons  $\Gamma_N^* = \partial D_N^+$ ,  $D_N^+ = \delta_0 \setminus (\cup_{n=1}^N \cup_{j=0}^{2^{[n\beta]}-1} \delta_{n,j})$ ,  $N = 1, 2, \dots$ , form decreasing approximation of  $\Gamma^*$ . The difference  $\Delta_N = \cup_{j=0}^{2^{[n\beta]}-1} \delta_{n,j}$  here is union of finite family of rectangles, and values of its perimeter and width are evident. For instance, if  $\varepsilon_n = O(\exp(-\alpha_n^{-1}))$ , then the series (6) converges for  $\nu > \frac{1}{2}$ , what is essentially better than  $\nu > \frac{\text{Dm } \Gamma^*}{2}$ .

4.3. A polygonal approximation can be constructed in terms of the Schauder series. Let  $z = \xi(x)$  be a continuous mapping of segment  $I = [0, 1]$  onto the curve  $\Gamma$ . We extend  $\xi(x)$  into the Schauder series  $\xi(x) = \sum_{j=0}^{\infty} a_j \Omega_j(x)$ , where  $\Omega_j(x)$  are the Schauder functions. These functions are piecewise-linear (see, for instance, [1]). Hence, a partial sum  $\xi_n(x) = \sum_{j=0}^n a_j \Omega_j(x)$  maps  $I$  onto a polygon  $\Gamma_n$ . But the Schauder series uniformly converges to  $\xi(x)$  (see [1]). Thus, the sequence  $\Gamma_1, \Gamma_2, \dots$  approximates the curve  $\Gamma$  (generally speaking, non-monotone one). Certain solvability conditions for the jump problem in terms of the Schauder coefficients are obtained in [9].

The research is supported by Russian Foundation for Basic Researches, grant 07-01-00166-a.

#### REFERENCES

- [1] Z. Ciesielski, Fractal functions and Schauder bases, *Comp. Math. Appl.*, 30, 283–291(1995).
- [2] E.M. Dynkin, Smoothness of the Cauchy type integral, *Zapiski nauchn. sem. Leningr. dep. mathem. inst. AN USSR*, 92, 115–133 (1979).
- [3] I. Feder, *Fractals*, Mir Publishers, Moscow, 1991.
- [4] F.D. Gakhov, *Boundary value problems*, Nauka publishers, Moscow, 1977.
- [5] B.A. Kats, The Riemann boundary value problem on closed Jordan curve, *Izv. VUZov, Mathematics*, 4, 68–80(1983).
- [6] B.A. Kats, The Riemann boundary value problem on non-smooth arcs and fractal dimensions, *Algebra and Analysis, St-Petersbourg*, 6, 147–171(1994).
- [7] B. A. Kats, The refined metric dimension with applications, *Computational Methods and Function Theory*, 7, 77–89(2007).
- [8] B. A. Kats, The Cauchy integral over Non-rectifiable Paths, *Contemporary Mathematics*, 455, 183–196(2008).
- [9] B. A. Kats and A. Yu. Pogodina, The jump problem and the Faber–Schauder series, *Izv. vuzov, Mathematics*, 1, 16–22(2007).
- [10] A.N. Kolmogorov, V.M. Tikhomirov,  $\varepsilon$ –entropy and capacity of set in functional spaces, *Uspekhi Math. Nauk*, 14, 3–86(1959).
- [11] R. Lesniewicz, W. Orlicz, On generalized variation II, *Studia Mathematica*, XLV, 71–109(1973).
- [12] N.I. Muskhelishvili, *Singular integral equations*, Nauka publishers, Moscow, 1962.
- [13] T. Salimov, A direct bound for the singular Cauchy integral along a closed curve, *Nauchn. Trudy Min. vyssh. i sr. spec. obraz. Azerb. SSR, Baku*, 5, 59–75(1979).
- [14] E.M. Stein, *Singular integrals and differential properties of functions*, Princeton University Press, Princeton, 1970.
- [15] L.C. Young, General inequalities for Stieltjes integrals and the convergence of Fourier series, *Math. Annalen*, 115, 581–612(1938).

# THE BOUNDARY VALUE PROBLEM FOR THE FOURTH-ORDER EQUATION WITH FRACTIONAL DERIVATIVES

D.AMANOV, N.M. KUZIBAEV

November 26, 2008

Institute of Mathematics and IT technologies , Tashkent, Uzbekistan.  
email:amanov\_d@rambler.ru.

**Keywords:** fractional derivatives (Caputo derivative, Riemann-Liouville derivative), 2-parabolic equation, equation of transverse vibration of elastic rods, Bessel inequality, orthonormal completeness in  $L_2(0, p)$  system, Mittag-Leffler function, Volterra integral equation of the second kind, Lipschitz conditions, piecewise continuous.

## 1 The statement of the problem.

In the domain  $\Omega = \{(x, t) : 0 < x < p, 0 < t < T\}$  let us consider the equation

$$u_{xxxx} + {}_C D_{0t}^\alpha u = f(x, t) \quad (1)$$

where  $1 < \alpha < 2$ ,  ${}_C D_{0t}^\alpha$  is Caputo fractional the  $\alpha$ th derivative operator with respect to  $t$  [1],[4].

If  $\alpha = 2$  the equation (1) changes to the well-known equation of transverse vibration of elastic rods

$$u_{xxxx} + u_{tt} = f(x, t),$$

And if  $\alpha = 1$  the equation (1) changes into 2parabolic equation

$$u_{xxxx} + u_t = f(x, t).$$

Caputo fractional derivative operator expresses as Riemann-Liouville fractional integral [1]

$${}_C D_{0t}^\alpha = I_{0t}^{2-\alpha} \frac{\partial^2 u}{\partial t^2} \quad (2)$$

**Problem.** In the domain  $\Omega = \{(x, t) : 0 < x < p, 0 < t < T\}$  find  $u(x, t)$  that is the solution of the equation (1) and is satisfying the following conditions

$$\frac{\partial^{2m} u(0, t)}{\partial x^{2m}} = \frac{\partial^{2m} u(p, t)}{\partial x^{2m}} = 0, m = 0, 1, 0 \leq t \leq T, \quad (3)$$



$$u(x, 0) = \varphi(x), u_t(x, 0) = \psi(x), 0 \leq x \leq p. \quad (4)$$

**Theorem.** If  $f(x, t) \in C^1[0, p], f(0, t) = f(p, t) = 0, \frac{\partial f(x, t)}{\partial x} \in Lip_\gamma[0, p]$  is uniformly in  $t, 0 < \gamma < 1. \varphi(x), \psi(x) \in C[0, p], \varphi'(x), \psi'(x)$  are piecewise continuous on  $[0, p], \varphi(0) = \varphi(p) = 0, \psi(0) = \psi(p) = 0$ , then there is a solution of the problem in class  $C_{x,t}^{2,1}(\bar{\Omega}) \cap C_{x,t}^{4,2}(\Omega)$ .

**Proof.** We research the regular solution of the problem in the form of

$$u(x, t) = \sum_{n=1}^{\infty} u_n(t) X_n(x), \quad (5)$$

here

$$X_n(x) = \sqrt{\frac{2}{p}} \sin \lambda_n x, \lambda_n = \frac{n\pi}{p}, n \in N.$$

Substitute (5) into the equation (1) we get

$${}_CD_{0,t}^\alpha u_n(t) + \lambda_n^4 u_n(t) = f_n(t) \quad (6)$$

where

$$f_n(t) = \int_0^p f(x, t) X_n(x) dx$$

Conditions (4) change to the following conditions

$$u_n(0) = \varphi_n, u'_n(0) = \psi_n$$

here

$$\varphi_n = \int_0^p \varphi(x) X_n(x) dx, \psi_n = \int_0^p \psi(x) X_n(x) dx.$$

Using (2) we reduce equation (6) to the form

$$I_{0,t}^{2-\alpha} u_n(t) + \lambda_n^4 u_n(t) = f_n(t)$$

Acting to both sides of this equation by the operator  $I_{0t}^\alpha$  we get Volterra integral equation of the second kind

$$u_n(t) + \frac{\lambda_n^4}{\Gamma(\alpha)} \int_0^t (t-\tau)^{\alpha-1} u_n(\tau) d\tau = \varphi_n + t\psi_n + I_{0t}^\alpha f_n(t)$$

Using the successive approximation method we find the unique solution

$$u_n(t) = \varphi_n E_{\alpha,1}(-\lambda_n^4 t^\alpha) + t\psi_n E_{\alpha,2}(-\lambda_n^4 t^\alpha) + \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\lambda_n^4 (t-\tau)^\alpha) f_n(\tau) d\tau$$

here

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(k\alpha + \beta)}$$

is Mittag-Leffler function[3].

Then the solution of the problem rewrites in the form of

$$\begin{aligned} u(x, t) = & \sum_{n=1}^{\infty} X_n(x) [\varphi_n E_{\alpha,1}(-\lambda_n^4 t^\alpha) + t \psi_n E_{\alpha,2}(-\lambda_n^4 t^\alpha)] + \\ & + \sum_{n=1}^{\infty} X_n(x) \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\lambda_n^4 (t-\tau)^\alpha) f_n(\tau) d\tau \end{aligned} \quad (7)$$

It is easy to check that

$$\begin{aligned} \lim_{t \rightarrow 0} u(x, t) &= \sum_{n=1}^{\infty} \varphi_n X_n(x) = \varphi(x), \\ \lim_{t \rightarrow 0} \frac{\partial u}{\partial t} &= \sum_{n=1}^{\infty} \psi_n X_n(x) = \psi(x). \end{aligned}$$

If we show the uniform convergence of the series

$$\begin{aligned} u_{xxxx} = & \sum_{n=1}^{\infty} \lambda_n^4 X_n(x) [\varphi_n E_{\alpha,1}(-\lambda_n^4 t^\alpha) + t \psi_n E_{\alpha,2}(-\lambda_n^4 t^\alpha)] + \\ & + \sum_{n=1}^{\infty} \lambda_n^4 X_n(x) \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\lambda_n^4 (t-\tau)^\alpha) f_n(\tau) d\tau \end{aligned} \quad (8)$$

then the series (7) and

$$\begin{aligned} {}_C D_{0t}^\alpha u = & \sum_{n=1}^{\infty} f_n(t) X_n(x) - \sum_{n=1}^{\infty} \lambda_n^4 X_n(x) [\varphi_n E_{\alpha,1}(-\lambda_n^4 t^\alpha) + t \psi_n E_{\alpha,2}(-\lambda_n^4 t^\alpha)] - \\ & - \sum_{n=1}^{\infty} \lambda_n^4 X_n(x) \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\lambda_n^4 (t-\tau)^\alpha) f_n(\tau) d\tau \end{aligned} \quad (9)$$

convergence uniformly, too.

The series (8) we represent as the sum  $I_1 + I_2 + I_3$ , where

$$\begin{aligned} I_1 &= \sum_{n=1}^{\infty} \lambda_n^4 \varphi_n X_n(x) E_{\alpha,1}(-\lambda_n^4 t^\alpha), \\ I_2 &= \sum_{n=1}^{\infty} \lambda_n^4 \psi_n X_n(x) t E_{\alpha,2}(-\lambda_n^4 t^\alpha), \\ I_3 &= \sum_{n=1}^{\infty} \lambda_n^4 X_n(x) \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\lambda_n^4 (t-\tau)^\alpha) f_n(\tau) d\tau. \end{aligned}$$

Now we need the following estimate

$$|E_{\alpha,\beta}(z)| \leq \frac{M}{1+|z|}, |z| > 0, M = \text{const} > 0 \quad (10)$$

Using (10) we show the convergence of the  $I_1$

$$\sum_{n=1}^{\infty} \lambda_n^4 |\varphi_n| |E_{\alpha,1}(-\lambda_n^4 t^\alpha)| \leq M \sum_{n=1}^{\infty} \frac{\lambda_n^4 |\varphi_n|}{1 + \lambda_n^4 t^\alpha}.$$

Let  $0 < t_0 < t \leq T$ . In terms of the theorem we have

$$\sum_{n=1}^{\infty} \frac{\lambda_n^4 |\varphi_n|}{1 + \lambda_n^4 t^\alpha} < \sum_{n=1}^{\infty} \frac{\lambda_n^4}{\lambda_n^4 t_0^\alpha} |\varphi_n| = \frac{1}{t_0^\alpha} \sum_{n=1}^{\infty} |\varphi_n| < \infty$$

Analogously for  $I_2$  we get

$$\sum_{n=1}^{\infty} \lambda_n^4 |\psi_n| t |E_{\alpha,2}(-\lambda_n^4 t^\alpha)| \leq \frac{M}{t_0^{\alpha-1}} \sum_{n=1}^{\infty} |\psi_n| < \infty.$$

And now we show the convergence of the  $I_3$ . It majorizes by the following series

$$\begin{aligned} & \sum_{n=1}^{\infty} \lambda_n^4 \int_0^t (t-\tau)^{\alpha-1} |E_{\alpha,\alpha}(-\lambda_n^4 (t-\tau)^\alpha)| |f_n(\tau)| d\tau \leq \\ & \leq CM \sum_{n=1}^{\infty} \frac{1}{\lambda_n^{1+\delta}} \int_0^t \frac{\lambda_n^4 (t-\tau)^\alpha}{1+\lambda_n^4 (t-\tau)^\alpha} d\tau = CM \sum_{n=1}^{\infty} \frac{(-1)}{\alpha \lambda_n^{1+\delta}} \int_0^t \frac{d(1+\lambda_n^4 (t-\tau)^\alpha)}{1+\lambda_n^4 (t-\tau)^\alpha} = \\ & = CM \sum_{n=1}^{\infty} \frac{1}{\lambda_n^{1+\delta}} \ln(2\lambda_n^4 t^\alpha) \leq CM \sum_{n=1}^{\infty} \frac{1}{\lambda_n^{1+\delta}} (\ln 2T^\alpha + \frac{4}{\varepsilon} \ln \lambda_n^\varepsilon), \varepsilon > 0 \end{aligned}$$

Let  $0 < \varepsilon < \delta$ . Then the above series converges uniformly.

Theorem is proved.

## References

1. Samko S.G., Kilbas A.A., Marichev O.U. Integrals and derivatives of fractional order and their applications. Minsk, Nauka and tehnika. 1987. 688p.
2. Mikhailov V.P. On potential of parabolic equations. Soviet math.dokl. > 129, N6, 1959, p.1226-1229.
3. Dzhrbashyan M.M. Integral transformations and representations of functions in complex domain. M., Nauka, 1966. 672p.
4. Gorenflo R., Luchko Y.F., Umarov S.R. on the Cauchy and multy-point problems for partial pseudo-differential equations of fractional order// Fract.Calc.Appl.2000,V.3.

# On Sparse Solutions of Underdetermined Linear Systems

Ming-Jun Lai

Department of Mathematics  
the University of Georgia  
Athens, GA 30602

January 15, 2009

## Abstract

We first explain the research problem of finding the sparse solution of underdetermined linear systems with some applications. Then we explain three different approaches how to solve the sparse solution: the  $\ell_1$  approach, the orthogonal greedy approach, and the  $\ell_q$  approach with  $0 < q \leq 1$ . We mainly survey recent results and present some new or simplified proofs. In particular, we give a good reason why the orthogonal greedy algorithm converges and why it can be used to find the sparse solution. About the restricted isometry property (RIP) of matrices, we provide an elementary proof to a known result that the probability that the random matrix with iid Gaussian variables possesses the RIP is strictly positive.

## 1 The Research Problem

Given a matrix  $\Phi$  of size  $m \times n$  with  $m \leq n$ , let

$$\mathcal{R}_k = \{\Phi \mathbf{x}, \mathbf{x} \in \mathbf{R}^n, \|\mathbf{x}\|_0 \leq k\}$$

be the range of  $\Phi$  of all the  $k$ -component vectors, where  $\|\mathbf{x}\|_0$  stands for the number of the nonzero components of  $\mathbf{x}$ .

Throughout this article,  $\Phi$  is assumed to be of full rank. For a vector  $\mathbf{y} \in \mathcal{R}_k$ , we solve the following minimization problem

$$\min\{\|\mathbf{x}\|_0, \quad \mathbf{x} \in \mathbb{R}^n, \Phi\mathbf{x} = \mathbf{y}\}. \quad (1)$$

The solution of the above problem is called the sparse solution of  $\mathbf{y} = \Phi\mathbf{x}$ .

It is clear that the above problem can be solved in a finite time. Indeed, write  $\Phi = [\phi_1, \phi_2, \dots, \phi_n]$  with  $\phi_i$  being a  $m \times 1$  vector. One can choose  $m$  columns, say  $A = [\phi_{i_1}, \dots, \phi_{i_m}]$  from  $\Phi$  to form a  $m \times m$  linear system:  $A\mathbf{z} = \mathbf{y}$ . If  $A$  is nonsingular, one can find a solution  $\mathbf{z}$ . By exhausting all  $m \times m$  nonsingular submatrices from  $\Phi$  and solving all such linear system of equations, one can see which solution has the smallest number of nonzero entries.

However, there could be  $C_m^n$  such nonsingular linear systems from  $\Phi\mathbf{x} = \mathbf{y}$  which need to be solved. For example, a rectangular matrix  $\Phi$  with entries  $(x_j)^i, i = 0, \dots, m, j = 1, \dots, n$  for distinct real numbers  $x_i$ 's. Any  $m \times m$  sub-matrix from  $\Phi$  is of full rank. The number  $C_m^n$  grows exponentially fast as  $m$  and  $n$  go to  $\infty$ . For example, when  $n = 2m$ ,  $C_m^n \approx 2^n$ . A common case  $m = 512$  and  $n = 1024$  needs to solve at least  $2^{512}$  linear systems of size  $512 \times 512$ . This is impossible to do within a hour using current available computer. That is, the above method to solve Eq. (1) needs non-polynomial time. Are there any other methods to solve the above problem? Before we answer this question, let us see why we want to solve the problem in the next section.

**Remark 1.1** When  $m = n$ ,  $\Phi\mathbf{x} = \mathbf{y}$  is a standard linear system and the solution is unique if  $\Phi$  is of full rank. We have already known the Gaussian elimination method can be used to solve such linear systems.

**Remark 1.2** When  $m > n$ , one may not be able to have  $\Phi\mathbf{x} = \mathbf{y}$ . Instead, one asks to find  $\mathbf{x}$  which minimizes the quantity  $\|\Phi\mathbf{x} - \mathbf{y}\|_2$ , where  $\|\cdot\|_2$  is the discrete  $\ell_2$  norm. This is a standard least squares problem. When  $\Phi$  is not full rank, one usually solves the following minimal norm solution using standard least squares methods. That is, find  $\mathbf{x}$  such that

$$\min\{\|\mathbf{x}\|_2, \quad \mathbf{x} \in S_A \subset \mathbb{R}^n, \}. \quad (2)$$

and

$$S_A := \{\mathbf{x} \in \mathbb{R}^n, \|\Phi\mathbf{x} - \mathbf{y}\|_2 = \min_{\mathbf{z}} \|\Phi\mathbf{z} - \mathbf{y}\|_2\}. \quad (3)$$

*The solution can be found by using the pseudo inverse or using the singular value decomposition.*

**Remark 1.3** *When each column of  $\Phi$  is normalized to be 1,  $\Phi$  is called a dictionary. When  $\Phi\Phi^T = I_m$  with identity matrix  $I_m$  of size  $m \times m$ ,  $\Phi$  is called a tight frame. We shall use these two concepts in later sections.*

## 2 Why do we find the sparse solutions?

In this section we give several reasons why we want to solve the sparse solution of underdetermined systems of linear equations.

### 2.1 Motivation: Signal and Image Compression

This is the most direct and natural application. Suppose that a signal or an image  $\mathbf{y}$  is represented by using a tight frame  $\Phi$  of size  $m \times n$  with  $m < n$ . We look for a sparse approximation  $\mathbf{x}$  satisfying

$$\min\{\|\mathbf{x}\|_0, \quad \mathbf{x} \in \mathbb{R}^n, \|\Phi\mathbf{x} - \mathbf{y}\| \leq \theta\}, \quad (4)$$

where  $\theta > 0$  is a tolerance. In particular, for lossless compression, i.e.,  $\theta = 0$ , the above (4) is the our research problem (1).

### 2.2 Motivation: Compressed Sensing

We are interested in economically recording information about a vector  $\mathbf{x}$  in  $\mathbb{R}^n$ . First of all, we allocate  $m$  nonadaptive questions to ask about  $\mathbf{x}$ . Each question takes the form of a linear functional applied to  $\mathbf{x}$ . Thus, the information we obtain from the questions is given by

$$\mathbf{y} = \Phi\mathbf{x},$$

where  $\Phi$  is a matrix of  $m \times n$ . In general,  $m$  is much smaller than  $n$  since  $\mathbf{x}$  is a compressible data vector. Let  $\Delta$  be a decoder that provides an approximation  $\mathbf{x}^*$  to  $\mathbf{x}$  using the information that  $\mathbf{y}$  holds. That is,  $\Delta\mathbf{y} = \mathbf{x}^* \approx \mathbf{x}$ . Typically, the mapping  $\Delta$  is nonlinear. The central question of compressed sensing is

to find a good set of questions and a good decoder  $(\Phi, \Delta)$  so that we can find a good approximation  $\mathbf{x}^*$  of  $\mathbf{x}$ . See, e.g. [Candés'06].

For example, when  $\mathbf{x} \in \mathbf{R}^n$  with  $\|\mathbf{x}\|_0 \leq k \ll n$ , one wants to know which components of  $\mathbf{x}$  are not zero and what values are. We can design some vectors  $\Phi$  to question  $\mathbf{x}$  by inner product. Thus, we get  $\mathbf{y} = \Phi\mathbf{x} \in \mathcal{R}_k$ . To find  $\mathbf{x}$ , we solve our research problem (1).

For another example, suppose that a data vector  $\mathbf{z}$  is a set of compressible data, i.e., there exists a vector  $\mathbf{x} \in \mathbf{R}^n$  with only  $k$  nonzero entries such that  $\mathbf{z} = A\mathbf{x}$  for an invertible matrix  $A$  of  $n \times n$ . Suppose that the questions can be represented in the form  $\mathbf{y} = C\mathbf{z}$  with  $m \times n$  matrix  $C$ . Since  $\mathbf{z} \in \mathcal{R}_k$ , i.e.,  $\mathbf{z} = A\mathbf{x}$ , we have

$$\mathbf{y} = C A \mathbf{x}. \quad (5)$$

Certainly it is necessary to question  $\mathbf{z}$   $m$  times with  $m > k$ . That is, we have  $k < m < n$ .

If  $C$  is chosen in the form  $\Phi A^{-1}$  for some rectangular matrix  $\Phi$  of size  $m \times n$ , we need to solve the following minimization problem in order to record the data  $\mathbf{z}$  economically.

$$\mathbf{y} = C \mathbf{z} = \Phi A^{-1} A \mathbf{x} = \Phi \mathbf{x}. \quad (6)$$

The problem is to find the sparse representation  $\mathbf{x}$  satisfying the above (6) which is the same as (1).

### 2.3 Motivation: Error Correcting Codes

Let  $\mathbf{z}$  be a vector encoded  $\mathbf{x}$  by a redundant linear system  $A$  of size  $m \times n$  with  $m > n$ . That is,  $\mathbf{z} = A\mathbf{x}$  is transmitted through a noisy channel. The channel corrupts some random entries of  $\mathbf{z}$ , resulting a new vector  $\mathbf{w} = \mathbf{z} + \mathbf{v}$ . Finding the vector  $\mathbf{v}$  is equivalent to correcting the errors.

To this end, we extend  $A$  to a square matrix  $B$  of size  $m \times m$  by adding  $A^\perp$ , i.e,  $B = [A; A^\perp]$ . Assume that  $A$  satisfies  $A^T A = I_n$ , where  $I_n$  is the identity matrix of  $n$ . Then we can choose  $A^\perp$  such that  $B B^T = I_m$  the identity matrix of size  $m$ . Clearly,

$$B^T \mathbf{w} = B^T \mathbf{z} + B^T \mathbf{v} = \begin{bmatrix} \mathbf{x} \\ 0 \end{bmatrix} + \begin{bmatrix} A^T \mathbf{v} \\ (A^\perp)^T \mathbf{v} \end{bmatrix}.$$

Let  $\mathbf{y} = (A^\perp)^T \mathbf{v}$  which is the last  $m - n$  entries of  $B^T \mathbf{w}$ . Since  $\mathbf{z}$  is in the codeword space  $V$  which is a linear span of columns of the matrix  $A$ ,  $(A^\perp)^T \mathbf{v}$

is not in the codeword space and is the only information about  $\mathbf{v}$  available to the receiver.

If the receiver is able to solve the minimization problem Eq.(1) with  $\Phi = (A^\perp)^T$ . That is, find the sparsest solution  $\mathbf{v}$  such that  $\mathbf{y} = (A^\perp)^T \mathbf{v}$ . Then we can get the correct  $\mathbf{x}$ . Thus, this error correcting problem is again equivalent to the sparsest solution problem Eq. (1). See [Candes and Tao'05] and [Candes, Romberg, Tao'06] for more detail.

## 2.4 Motivation: Cryptography

Although large prime numbers are currently used for secure data transmission, it is possible to use underdetermined systems of linear equations instead. The ideas can be described as follows. Suppose that we have a class of matrices  $\Phi$  of size  $m \times n$  with  $m < n$  which admit a computationally efficient algorithm for solving the minimization Eq.(1) for any given  $\mathbf{y}$  which is in the range  $\mathcal{R}_k$ . Let  $\Psi$  be an invertible random matrix of size  $m \times m$  and  $A = \Psi\Phi$ . Suppose that a receiver wants to get a secret data vector  $\mathbf{x}$  from a customer, e.g., a vector consisting of credit card number, expiration date, and the name on the credit card. The receiver sends to the customer the matrix  $A$  in a public channel. After receiving  $A$  the customer computes  $\mathbf{z} = A\mathbf{x}$  and sends  $\mathbf{z}$  to the receiver in a public channel. As we mentioned above, finding the sparse solution  $\mathbf{x}$  from  $\mathbf{z}$  using matrix  $A$  is non-polynomial time. With overwhelming probability, such  $\mathbf{x}$  can not be found by other parties.

However, the receiver is able to get  $\mathbf{x}$  by solving  $\mathbf{y} = \Psi^{-1}\mathbf{z} = \Phi\mathbf{x}$  which is our research problem Eq. (1). By changing  $\Psi$  frequently enough, the receiver is able to get the secured data every time while the hacker is impossible to decode the data.

## 2.5 Motivation: Recovery of Loss Data

Let  $\mathbf{z}$  be an image and  $\tilde{\mathbf{z}}$  be a partial image of  $\mathbf{z}$ . That is,  $\mathbf{z}$  loses some data to become  $\tilde{\mathbf{z}}$ . Suppose that we know the location where the data are lost. We would like to recover the original image from the partial image  $\tilde{\mathbf{z}}$ . Let  $\Phi$  be a tight wavelet frame such that  $\mathbf{x} = \Phi\mathbf{z}$  is the most sparse representation for  $\mathbf{z}$ . Let  $\Psi$  be the residual matrix from  $\Phi$  by dropping off the columns corresponding to the unavailable entries, i.e., the missing data locations. Note that  $\Phi^T\Phi = I_m$  and hence,  $\Psi^T\Psi = I_\ell$  with  $\ell < m$ .



It follows that  $\Psi^T \mathbf{x} = \Psi^T \Phi \mathbf{z} = \tilde{\mathbf{z}}$  by the orthonormality of columns of  $\Phi$ . Thus, we need to find the sparsest solution  $\mathbf{x}$  from the given  $\tilde{\mathbf{z}}$  such that  $\tilde{\mathbf{z}} = \Psi^T \mathbf{x}$  which is exactly the same problem as our research problem(1). Once we have  $\mathbf{x}$ , we can find  $\mathbf{z}$  which is  $\mathbf{z} = \Phi \mathbf{x}$ . See [Aharon, Elad, Bruckstein'06] for numerical experiments.

### 3 The $\ell_1$ Approach

Although the problem in Eq. (1) needs a non-polynomial time to solve (cf. [Natarajan95]) in general, it can be much more effectively solved by using many other methods, e.g.,  $\ell^1$  minimization approach, reweighted  $\ell^1$  method, OGA(orthogonal greedy algorithm), and the  $\ell_q$  approach. Let us review these approaches in the following subsections and following sections.

The  $\ell_1$  minimization problem is the following

$$\min\{\|\mathbf{x}\|_1, \quad \mathbf{x} \in \mathbb{R}^n, \Phi \mathbf{x} = \mathbf{y}\}, \quad (7)$$

where  $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$  for  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ . The solution  $\Delta_1 \Phi \mathbf{x}$  is called the  $\ell_1$  solution of  $\mathbf{y} = \Phi \mathbf{x}$ . Since the  $\ell_1$  minimization problem is equivalent to the linear programming, this converts the problem into a tractable computational problem. (See [Lai and Wenston'04] for a justification of the equivalence and a computational algorithm for  $\ell_1$  minimization.) A matlab  $\ell_1$  minimization program is available on-line.

But one has to study when the (P1) solution (the solution of Eq. (7)) is also the (P0) solution (the solution of Eq. (1)). There are two concepts: mutual coherence(MC) and restricted isometric property (RIP) of the matrix  $\Phi$  to help describe the situation.

#### 3.1 The Mutual Coherence

Let us begin with the *spark* of matrix  $A$ , the smallest possible number  $\sigma$  such that there exists  $\sigma$  columns from  $A$  that are linearly dependent. It is clear that  $\sigma(A) \leq \text{rank}(A) + 1$ . The following theorem is belong to [Donoho and Elad'03].

**Theorem 3.1** *A representation  $\mathbf{y} = \Phi \mathbf{x}$  is necessarily the sparsest possible if  $\|\mathbf{x}\|_0 < \text{spark}(\Phi)/2$ .*

**Proof.** Suppose that there are two sparse solutions  $\mathbf{x}^1$  and  $\mathbf{x}^2$  with  $\|\mathbf{x}^1\|_0 \leq k$  and  $\|\mathbf{x}^2\|_0 \leq k$  solving  $\mathbf{y} = \Phi\mathbf{x}$ . Then  $\Phi(\mathbf{x}^1 - \mathbf{x}^2) = 0$ . So  $\|(\mathbf{x}^1 - \mathbf{x}^2)\|_0 \leq 2k$  but,  $\|(\mathbf{x}^1 - \mathbf{x}^2)\|_0 \geq \text{spark}(\Phi)$ . It follows that  $k \geq \text{spark}(\Phi)/2$ . Hence, when  $k < \text{spark}(\Phi)/2$ , the sparsest solution is unique. That is, if one find a solution  $\mathbf{x}$  of  $\Phi\mathbf{x} = \mathbf{y}$  with  $\|\mathbf{x}\|_0 < \text{spark}(\Phi)/2$ , then  $\mathbf{x}$  is the sparse solution. ■

Next we introduce the concept of mutual coherence of matrix  $\Phi$ . Assume that each column of  $\Phi$  is normalized. That is,  $\Phi$  is a dictionary. Let  $G = \Phi^T\Phi$  which is a square matrix of size  $n \times n$ . Write  $G = (g_{ij})_{1 \leq i, j \leq n}$ , the mutual coherence of  $\Phi$  is

$$M = M(\Phi) = \max_{\substack{1 \leq i, j \leq n \\ i \neq j}} |g_{ij}|.$$

Clearly,  $M \leq 1$ . We would like to have matrix  $\Phi$  such that its mutual coherence  $M$  is as small as possible.

However,  $M(\Phi)$  can not be too small. We have

**Lemma 3.2** *If  $n \geq 2m$ , then  $M(\Phi) \geq (2m)^{-1/2}$ .*

**Proof.** Indeed, let  $\lambda_i, i = 1, \dots, n$  be eigenvalues of  $G$ . Since  $G$  is positive semi-definite, all  $\lambda_i \geq 0$ . Since the rank of  $G$  is equal to  $m$ , only  $m$  nonzero  $\lambda_i$ . Since  $\sum_i \lambda_i$  is equal to the trace of  $G$  which is  $n$  since  $g_{ii} = 1$ . That is,

$$n = \sum_i \lambda_i \leq \sqrt{m} \sqrt{\sum_i \lambda_i^2}. \quad (8)$$

On the other hand, using a property of the Frobenius norm of  $G$ , we have

$$\sum_i \lambda_i^2 = \|G\|_F^2 = \sum_{1 \leq i, j \leq n} (g_{ij})^2. \quad (9)$$

It follows from Eq. (8) and (9) that

$$(n^2 - n)M(\Phi)^2 + n \geq \sum_{1 \leq i, j \leq n} (g_{ij})^2 \geq \frac{n^2}{m}$$

That is,  $M(\Phi) \geq \sqrt{\frac{n-m}{m(n-1)}}$ . In particular, when  $n \geq 2m$ , we have  $M(\Phi) \geq (2m)^{-1/2}$ . That is,  $M(\Phi) \in ((2m)^{-1/2}, 1]$ . ■

With  $M(\Phi)$ , we can prove the following (cf. [Donoho and Elad03])

**Theorem 3.3** *Let  $\text{Spark}(\Phi)$  be the spark of  $\Phi$  and  $M(\Phi)$  be the coherence of  $\Phi$ . Then*

$$\text{Spark}(\Phi) > 1/M(\Phi).$$

Next we need the following lemma.

**Lemma 3.4** *Let  $k < 1/M + 1$ . For any  $S \subset \{1, \dots, n\}$  with  $\#(S) \leq k$  and  $\Phi_S$  be the matrix consisting of the  $k$  columns of  $\Phi$  with column indices in  $S$ . Then the  $k$ th singular value of  $\Phi_S$  is bounded below by  $(1 - M(k - 1))^{1/2}$  and above by  $(1 + M(k - 1))^{1/2}$ .*

**Proof.** For any vector  $\mathbf{v} \in \mathbf{R}^n$  with support on  $S$ , we have

$$\mathbf{v}^T G \mathbf{v} = \mathbf{v}_S \Phi_S^T \Phi_S \mathbf{v}_S = \|\mathbf{v}\|^2 + \sum_{\substack{i \neq j \\ i, j \in S}} v_i g_{ij} v_j.$$

Since

$$\begin{aligned} \left\| \sum_{\substack{i \neq j \\ i, j \in S}} v_i g_{ij} v_j \right\| &\leq M \sum_{\substack{i \neq j \\ i, j \in S}} |v_i v_j| \\ &\leq M \left( \sum_{i, j \in S} |v_i v_j| - \|\mathbf{v}\|_2^2 \right) \\ &\leq M \|\mathbf{v}\|_2^2 (k - 1), \end{aligned}$$

we have

$$\mathbf{v}_S \Phi_S^T \Phi_S \mathbf{v}_S \geq \|\mathbf{v}\|_2^2 - M(k - 1) \|\mathbf{v}\|_2^2 = (1 - M(k - 1)) \|\mathbf{v}\|_2^2.$$

Similarly,

$$\begin{aligned} \mathbf{v}^T G \mathbf{v} &= \mathbf{v}_S \Phi_S^T \Phi_S \mathbf{v}_S = \|\mathbf{v}\|^2 + \sum_{\substack{i \neq j \\ i, j \in S}} v_i g_{ij} v_j \\ &\leq \|\mathbf{v}\|_2^2 + M(k - 1) \|\mathbf{v}\|_2^2 = (1 + M(k - 1)) \|\mathbf{v}\|_2^2. \end{aligned}$$

These complete the proof. ■

We first show that if  $k < (1 + 1/M)/2$  and for any  $\mathbf{y} \in \mathcal{R}_k$ , the sparse solution of Eq. (1) is unique.

**Lemma 3.5** *Suppose  $k < (1 + 1/M)/2$ . For any  $\mathbf{y} \in \mathcal{R}_k$ , the sparse solution of Eq. (1) is unique.*

**Proof.** Let  $\mathbf{y} = \Phi \mathbf{x}_0 \in \mathcal{R}_k$ . Let  $\mathbf{x}_1$  be a solution of Eq. (1) with  $\|\mathbf{x}_1\|_0 \leq k$ . Then  $\|\mathbf{x}_0 - \mathbf{x}_1\|_0 \leq 2k$ . By Lemma 3.4, we have

$$(1 - M(2k - 1))\|\mathbf{x}_0 - \mathbf{x}_1\|_2^2 \leq \|\Phi(\mathbf{x}_0 - \mathbf{x}_1)\|_2^2 = 0.$$

Since  $1 - M(2k - 1) \neq 0$ , it follows that  $\mathbf{x}_0 = \mathbf{x}_1$ . ■

In order to fully understand the computation of Eq. (7), we next introduce a variance of Eq. (7): solve the following minimization problem

$$\min\{\|\mathbf{x}\|_1, \quad \mathbf{x} \in \mathbb{R}^n, \|\Phi \mathbf{x} - \mathbf{y}\|_2 \leq \delta\}, \quad (10)$$

where  $\mathbf{y} = \Phi \mathbf{x}_0 + z$  with  $\|z\|_2 \leq \epsilon < \delta$  and  $\mathbf{x}_0 \in \mathbb{R}^n$  is a vector with  $k$  nonzero entries, that is,  $\Phi \mathbf{x}_0 \in \mathcal{R}_k$ . Hence, we consider the case that  $\mathbf{y}$  has some measurement error and compute a solution  $\mathbf{x}$  within accuracy  $\delta$ .

**Theorem 3.6** *Let  $M$  be the mutual coherence of  $\Phi$ . Suppose that*

$$k < (1/M + 1)/4.$$

*For any  $\mathbf{x}_0$  with  $\|\mathbf{x}_0\|_0 \leq k$ , let  $\hat{\mathbf{x}}_{\epsilon, \delta}$  be the solution of Eq. (10). Then*

$$\|\hat{\mathbf{x}}_{\epsilon, \delta} - \mathbf{x}_0\|_2^2 \leq \frac{(\epsilon + \delta)^2}{1 - M(4k - 1)}.$$

**Proof.** Write  $w = \hat{\mathbf{x}}_{\epsilon, \delta} - \mathbf{x}_0$ . Clearly,  $\|\hat{\mathbf{x}}_{\epsilon, \delta}\|_1 = \|w + \mathbf{x}_0\|_1 \leq \|\mathbf{x}_0\|_1$  when computing the  $\ell_1$  minimization. Let  $S \subset \{1, 2, \dots, n\}$  be the index set where  $\mathbf{x}_0$  is supported. Since  $\|w + \mathbf{x}_0\|_1 \geq \|\mathbf{x}_0\|_1 - \sum_{i \in S} |w_i| + \sum_{i \in \hat{S}} |w_i|$ , we have  $\sum_{i \in \hat{S}} |w_i| \leq \sum_{i \in S} |w_i|$  or

$$\|w\|_1 \leq 2 \sum_{i \in S} |w_i| \leq 2\sqrt{k}\|w\|_2, \quad (11)$$

where  $\hat{S}$  denotes the complement set of  $S$  in  $\{1, 2, \dots, n\}$ .

On the other hand,  $\|\Phi \hat{\mathbf{x}}_{\epsilon, \delta} - \mathbf{y}\|_2 \leq \delta$  and  $\mathbf{y} = \Phi \mathbf{x}_0 + z$  imply that  $\|\Phi w + z\|_2 \leq \delta$ . That is,  $\|\Phi w\|_2 \leq \|\Phi w + z\|_2 + \epsilon \leq \delta + \epsilon$ .

Finally,  $\|\Phi w\|_2^2 = \|w\|_2^2 + w^T(G - I)w \geq \|w\|_2^2 - M(\|w\|_1^2 - \|w\|_2^2) \geq (1 + M)\|w\|_2^2 - M4k\|w\|_2^2$  by the estimate (11) above. It follows that

$$\|w\|_2^2 \leq \frac{1}{1 + M - 4Mk} \|\Phi w\|_2^2 \leq \frac{(\epsilon + \delta)^2}{1 - M(4k - 1)}$$

by the estimate in the previous paragraph. This concludes the result in this theorem. ■

Next we look at the  $(\epsilon, \delta)$  variance of the (P0) problem: to solve the following minimization problem

$$\min\{\|\mathbf{x}\|_0, \quad \mathbf{x} \in \mathbb{R}^n, \|\Phi\mathbf{x} - \mathbf{y}\|_2 \leq \delta\}, \quad (12)$$

where  $\mathbf{y} = \Phi\mathbf{x}_0 + \mathbf{z}$  with  $\|\mathbf{z}\|_2 \leq \epsilon$  and  $\Phi\mathbf{x}_0 \in \mathcal{R}_k$ . Then we can prove

**Theorem 3.7** *Let  $M$  be the mutual coherence of  $\Phi$ . Suppose that*

$$k < (1/M + 1)/2.$$

*For any  $\mathbf{x}_0$  with  $\|\mathbf{x}_0\|_0 \leq k$ , let  $\tilde{\mathbf{x}}_{\epsilon, \delta}$  be the solution of Eq. (12). Then*

$$\|\tilde{\mathbf{x}}_{\epsilon, \delta} - \mathbf{x}_0\|_2^2 \leq \frac{(\epsilon + \delta)^2}{1 - M(2k - 1)}.$$

**Proof.** By Lemma 3.4, we have

$$\begin{aligned} \|\tilde{\mathbf{x}}_{\epsilon, \delta} - \mathbf{x}_0\|_2^2 &\leq \frac{1}{1 - M(2k - 1)} \|\Phi(\tilde{\mathbf{x}}_{\epsilon, \delta} - \mathbf{x}_0)\|_2^2 \\ &= \frac{1}{1 - M(2k - 1)} \|\Phi\tilde{\mathbf{x}}_{\epsilon, \delta} - \mathbf{y} + \mathbf{z}\|_2^2 \leq \frac{(\epsilon + \delta)^2}{1 - M(2k - 1)}. \end{aligned}$$

This completes the proof. ■

Both theorems above were proved in [Donoho, Elad and Temlyakov'06]. In particular, the proof of Theorem 3.7 above is a much simplified version of the one in [Donoho, Elad and Temlyakov'06]. It is easy to see that there is a gap between the requirements of  $k$ . That is, one is to require  $k < (1 + 1/M)/4$  by any  $\ell_1$  method and the other is to require  $k < (1 + 1/M)/2$  by an  $\ell_0$  method. Thus, the  $\ell_1$  method is not optimal yet. It is interesting to know how we can increase  $k$  when using the  $\ell_1$  method.

## 3.2 RIP

Another approach is to use the so-called Restricted Isometry Property (RIP) of the matrix  $\Phi$ . Letting  $0 < k < m$  be an integer and  $A_T$  be a submatrix of  $A$

which consists of columns of  $A$  whose column indices are in  $T \subset \{1, 2, \dots, n\}$ , the  $k$  restricted isometry constant  $\delta_k$  of  $A$  is the smallest quantity such that

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 \leq \|A_T \mathbf{x}\|_2^2 \leq (1 + \delta_k)\|\mathbf{x}\|_2^2 \quad (13)$$

for all subset  $T$  with  $\#(T) \leq k$ . If a matrix  $A$  has such a constant  $\delta_k > 0$  for some  $k$ ,  $A$  possesses RIP. With this concept, it is easy to see that if  $\delta_{2k} < 1$ , then the solution of Eq. (1) is unique. Indeed, if there were two solutions  $\mathbf{x}^1$  and  $\mathbf{x}^2$  such that

$$\Phi(\mathbf{x}^1 - \mathbf{x}^2) = 0,$$

then we choose the index set  $T$  which contains the indices of the nonzero entries of  $\mathbf{x}^1 - \mathbf{x}^2$  and see that  $\#(T) \leq 2k$  which implies

$$(1 - \delta_{2k})\|\mathbf{x}^1 - \mathbf{x}^2\|_2^2 \leq \|\Phi_T(\mathbf{x}^1 - \mathbf{x}^2)\|_2^2 = 0.$$

It follows that  $\|\mathbf{x}^1 - \mathbf{x}^2\|_2 = 0$  when  $\delta_{2k} < 1$ . That is, the solution is unique. Furthermore,

**Theorem 3.8** ([Candes, Romberg, and Tao'06]) *Suppose that  $k \geq 1$  such that*

$$\delta_{3k} + 3\delta_{4k} < 2$$

*and let  $\mathbf{x} \in \mathbf{R}^n$  be a vector with  $\|\mathbf{x}\|_0 \leq k$ . Then for  $\mathbf{y} = \Phi\mathbf{x}$ , the solution of Eq. (7) is unique and equal to  $\mathbf{x}$ .*

This result is recently simplified slightly in the following way:

**Theorem 3.9** ([Candes'08]) *Suppose that  $k \geq 1$  such that*

$$\delta_{2k} < \sqrt{2} - 1.$$

*Let  $\mathbf{x} \in \mathbf{R}^n$  be a vector with  $\|\mathbf{x}\|_0 \leq k$ . Then for  $\mathbf{y} = A\mathbf{x}$ , the solution of Eq. (7) is unique and equal to  $\mathbf{x}$ . In fact*

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \frac{2(1 + \rho)}{1 - \rho} \|A\mathbf{x} - A\mathbf{x}^*\|_2 + \frac{2}{1 - \rho} \|\mathbf{x} - \mathbf{x}_T^*\|_1.$$

*where  $\mathbf{x}^*$  is the (P1) solution (the solution of Eq. (7) and  $\mathbf{x}_T^*$  is the vector of the  $k$  largest components of  $\mathbf{x}^*$ . Here  $\rho = \frac{\delta_{2k}}{\sqrt{2}-1}$ .*

As we have already known that  $\delta_{2k} < \sqrt{2} - 1 < 1$  which implies that the sparse solution is unique. The above result mainly explains that the (P1) solution (the solution of Eq. (7)) is equal to the (P0) solution (the solution of Eq. (1)).

The results are consequences of the following Theorem 5.2 and hence we omit the proofs of the above two theorems here.

Let us discuss what kind of matrices  $\Phi$  satisfies the RIP. So far there is no explicit construction of matrices of any size which possess the RIP. Instead, there are a couple of constructions based on random matrices which satisfy the RIP with overwhelming probability. In [Candés, Romberg, and Tao'06], the following results were proved using the measure concentration technique (cf. [Ledoux'01]).

**Theorem 3.10** *Suppose that  $A = [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$  be a matrix with entries  $a_{ij}$  being iid Gaussian random variables with mean zero and variance  $1/\sqrt{m}$ . Then the probability*

$$\mathcal{P} \left( \left| \|\Phi \mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \leq \epsilon \|\mathbf{x}\|_2^2 \right) \geq 1 - \binom{n}{k} (1 + 2/\epsilon)^k e^{-m\epsilon^2/c}. \quad (14)$$

for any vector  $\mathbf{x} \in \mathbf{R}^n$  with  $\|\mathbf{x}\|_0 = k$ , where  $c > 2$  is a constant and  $\|\mathbf{x}\|_0$  denotes the number of nonzero entries of vector  $\mathbf{x}$ .

Once we choose  $k < m$  such that  $\binom{n}{k} (1 + 2/\epsilon)^k e^{-m\epsilon^2/c} < 1$  small enough, we will have a good probability to have a matrix satisfying the RIP. Indeed, since  $\binom{n}{k} \leq (n/e)^k$ ,

$$\binom{n}{k} (1 + 2/\epsilon)^k e^{-m\epsilon^2/c} \leq e^{-m\epsilon^2/c + k \ln(n/e) + k \ln(1+2/\epsilon)}.$$

As long as  $m > kc \ln(n(1 + 2/\epsilon)/e)/\epsilon^2$ , we have

$$\mathcal{P} \left( \left| \|\Phi \mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \leq \epsilon \|\mathbf{x}\|_2^2 \right) > 0.$$

That is, a matrix with RIP can be found with positive probability.

Theorem 3.10 can be simply proved based on the following theorem (cf. [Baranuik, Daveport, DeVore, Wakin'08]).

**Theorem 3.11** *Suppose that  $A = [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$  be a matrix with entries  $a_{ij}$  being iid Gaussian random variables with mean zero and variance  $1/\sqrt{m}$ . Then for any  $\epsilon > 0$ , the probability*

$$\mathcal{P}(|\|A\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2| < \epsilon \|\mathbf{x}\|_2^2) \geq 1 - 2 \exp(-\frac{\epsilon^2 m}{c}), \quad (15)$$

where  $c$  is a positive constant independent of  $\epsilon$  and  $\|\mathbf{x}\|_2$  for any  $\mathbf{x} \in \mathbb{R}^n$ .

In general, there are many other random matrices satisfying the above probability estimate. Typically, matrices with sub-Gaussian random variables possess the RIP. See [Mendelson, Pajor, and Tomczak-Jaegermann'07, '08]. In addition to the measure concentration approach, there are several other ideas to prove the results in the above theorem. For example, [Pisier'86] and [Lai'08]. We refer to [Lai'08] for an elementary proof of Theorems 3.11 and 3.10 and similar theorems for sub-Gaussian random matrices. For convenience, we borrow the proof of Theorem 3.11 from [Lai'08] and present it in the Appendix for interested reader.

### 3.3 The re-weighted $\ell_1$ Method

The re-weighted  $\ell_1$  minimization is the following iterations:

- (1) for  $k = 0$ , solve the standard  $\ell_1$  problem:

$$\min\{\|\mathbf{x}\|_1, \quad \mathbf{x} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{y}\}, \quad (16)$$

- (2) for  $k > 0$ , find  $\mathbf{x}^{(k)}$  which solves the following weighted  $\ell_1$  problem:

$$\min \sum_{i=1}^n \frac{|x_i|}{w_i}, \quad \mathbf{x} \in \mathbb{R}^n, A\mathbf{x} = \mathbf{y}, \quad (17)$$

with  $w_i = |x_i^{(k-1)}| + \epsilon$  for  $k = 1, 2, 3, \dots, n$ .

This method is introduced in [Candés, Watkin, and Boyd'07]. The researchers gave some heuristic reasons that the algorithm above converges much faster than the standard  $\ell_1$  method. It is still interesting why the method works better in theory.



## 4 The OGA Approach

There are many versions of the Optimal Greedy Algorithm(OGA) available in the literature. See [Temlyakov'00], [Temlyakov'03], [Tropp'04], and [Petukhov'06]. We mainly explain the optimal greedy algorithm (OGA) proposed by A. Petukhov in 2006 when  $\Phi$  is obtained from a tight wavelet frame. That is,  $\Phi$  is a matrix whose columns are frame components  $\phi_i, i = 1, \dots, n$  satisfying  $\Phi\Phi^T = I_m$ , where  $I_m$  is the identity matrix of size  $m \times m$ . It has two distinct advantages: (1) Iterative steps for the least squares solution and (2) more than one terms are chosen in each iteration.

Let  $\Lambda$  be an index set which is a subset of  $\{1, 2, \dots, n\}$  and  $\tilde{\Lambda}$  be the complement of  $\Lambda$  in  $\{1, 2, \dots, n\}$ . Also let  $P_\Lambda$  be the diagonal matrix of size  $n \times n$  with entries to be 1 if the index is in  $\Lambda$  and 0 otherwise.

Suppose that we have a fixed index set  $\Lambda$ . We first introduce a computationally efficient algorithm for finding coefficients of the linear combination  $f_\Lambda = \sum_{i \in \Lambda} a_i \phi_i$  which is the least squares approximation of  $f$ , i.e.,  $\|f - f_\Lambda\|_2 = \min\{\|f - g\|_2, g \in S_\Lambda\}$  where  $f \in \mathbf{R}^m$  is a given vector in  $\mathbf{R}^m$  and  $S_\Lambda$  is the span of  $\phi_i, i \in \Lambda$ . In general,  $f_\Lambda$  can be computed directly by inverting a Gram matrix  $[\langle \phi_i, \phi_j \rangle]_{i,j \in \Lambda}$ . When  $m$  is large, it is more efficient to use the following algorithm to find an approximation of  $f_\Lambda$ .

**Algorithm LSA (least squares approximation):** Set  $k = 0, g^0 = f, f^0 = 0$ . For  $k \geq 1$ , let  $g^k = g^{k-1} - \Phi P_\Lambda \Phi^T g^{k-1}$  and  $f^k = f^{k-1} + \Phi P_\Lambda \Phi^T g^{k-1}$ . Stop the iterations when  $g^k - g^{k-1}$  is very small.

We have the following

**Theorem 4.1** *The sequence  $f^k$  converges to  $f_\Lambda$  in the following sense:*

$$\|f^k - f_\Lambda\|_2 \leq (1 - \gamma^2)^{k/2} \|f_\Lambda\|_2,$$

where  $\gamma$  is the least non-zero singular value of the matrix  $\Phi_\Lambda$ .

**Proof.** We rewrite  $g^k$  as  $g^k = g_\Lambda^k + \tilde{g}^k$ , where  $g_\Lambda^k$  is the best approximation of  $g^k$  using the span of columns from  $\Phi_\Lambda$ . Clearly,  $g_\Lambda^0 = f_\Lambda$ . For  $k \geq 1$ ,  $g_\Lambda^k = g_\Lambda^{k-1} - \Phi P_\Lambda \Phi^T g_\Lambda^{k-1}$  since the best approximation operator in  $\mathbf{R}^m$  is a linear operator. Similar for  $f_\Lambda^k = f_\Lambda^{k-1} + \Phi P_\Lambda \Phi^T g_\Lambda^{k-1}$ . Note that  $f_\Lambda^k = f_\Lambda^k$  for all  $k \geq 0$ . We have

$$f^k + g_\Lambda^k = f_\Lambda^k + g_\Lambda^k = f_\Lambda^{k-1} + g_\Lambda^{k-1} = \dots = f_\Lambda^0 + g_\Lambda^0 = f_\Lambda.$$

It follows that

$$\|f^k - f_\Lambda\|_2 = \|g_\Lambda^k\|_2 = \|(I - \Phi P_\Lambda \Phi^T)g_\Lambda^{k-1}\|_2.$$

Note that  $I - \Phi P_\Lambda \Phi^T = \Phi(I - P_\Lambda)\Phi^T$  and hence

$$\|I - \Phi P_\Lambda \Phi^T\|_2 \leq \|\Phi(I - P_\Lambda)\Phi^T\|_2 \leq (1 - \gamma^2)^{1/2}.$$

Therefore,

$$\begin{aligned} \|f^k - f_\Lambda\|_2 &= \|g_\Lambda^k\|_2 \leq (1 - \gamma^2)^{1/2} \|g_\Lambda^{k-1}\|_2 \\ &\leq \cdots \leq (1 - \gamma^2)^{k/2} \|g_\Lambda^0\|_2 = (1 - \gamma^2)^{k/2} \|f_\Lambda\|_2. \end{aligned}$$

This completes the proof. ■

We are now ready to present the Petukhov version of orthogonal greedy algorithm (OGA).

**Algorithm OGA:** Set  $\Lambda_0 = \emptyset$ ,  $g^0 = f$ ,  $f^0 = 0$ . Choose a threshold  $r \in (0, 1]$  and a precision  $\epsilon > 0$ ;

Step 1. For  $k \geq 1$ , find  $M_k = \max_{i \notin \Lambda_{k-1}} |\langle g^{k-1}, \phi_i / \|\phi_i\| \rangle|$ ; and Let  $\Lambda_k = \Lambda_{k-1} \cup \{i, |\langle g^{k-1}, \phi_i / \|\phi_i\| \rangle| \geq r M_k\}$ ;

Step 2. Apply Algorithm LSA above over  $\Lambda_k$  to approximate  $g^{k-1}$  to find  $f_{\Lambda_k}$  and  $g_{\Lambda_k}$ . Update  $f^k = f^{k-1} + f_{\Lambda_k}$  and  $g^k = g^{k-1} - f_{\Lambda_k}$ .

Step 3. If  $\|f - f^k\|_2 \leq \epsilon$ , we stop the algorithm. Otherwise we advance  $k$  to  $k + 1$  and go to Step 1.

There is lack of theory to justify why the above OGA is convergent in the original paper [Petukhov'06] and in the literature so far. We now present an analysis of the convergence of the above OGA.

**Theorem 4.2** *Suppose that  $\Phi$  of size  $n \times N$  has the RIP for order  $k$  with  $1 \leq k \leq n$ . Then the above OGA converges.*

**Proof.** Without loss of generality we may assume that  $\Lambda_m = \{1, 2, \dots, n_m\}$  for some  $n_m < n$ , where  $m = 1, 2, \dots$ . Let

$$G_m = [\langle \phi_i, \phi_j \rangle]_{1 \leq i, j \leq n_m}$$

be the Grammian matrix. Define

$$a_m \leq \|G_m\|_2 \leq b_m$$

to be the smallest and largest eigenvalues of symmetric  $G_m$ . The RIP of  $\Phi$  for integer  $n_m$  implies that  $a_m > 0$  for  $m = 1, 2, \dots, m_0$  with  $n_{m_0} = n$ .

We first observe that the best approximation  $f_{\Lambda_m} = \Phi_m^{-1} [\langle f, \phi_i \rangle]_{1 \leq i \leq n_m}^T$ .

Then due to the result in Theorem 4.1, let us for simplicity, assume that  $f_{\Lambda_m}$  is the best approximation of  $R_{m-1}(f)$ . We next note that for  $i \in \Lambda_m \setminus \Lambda_{m-1}$ ,

$$|\langle R_{m-1}(f), \phi_i / \|\phi_i\| \rangle| \geq r M_m$$

with

$$\begin{aligned} M_m &= \max_{i \notin \Lambda_{m-1}} |\langle R_{m-1}(f), \phi_i / \|\phi_i\| \rangle| = \max_{i=1, \dots, n} |\langle R_{m-1}(f), \phi_i / \|\phi_i\| \rangle| \\ &\geq \left| \sum_{i=1}^n \alpha_i \langle R_{m-1}(f), \phi_i \rangle \right| \end{aligned}$$

for any  $\alpha_i$  such that  $\sum_{i=1}^n |\alpha_i| \leq 1$ . Assume that  $f = \sum_{i=1}^n c_i \phi_i$  with  $\sum_{i=1}^n |c_i| \leq 1$  (with appropriate normalization). It follows that

$$M_m \geq |\langle R_{m-1}(f), f \rangle| = \|R_{m-1}(f)\|^2.$$

Hence we have

$$\|R_m(f)\|^2 = \langle R_{m-1}(f) - f_{\Lambda_m}, R_{m-1}(f) - f_{\Lambda_m} \rangle = \|R_{m-1}(f)\|^2 - \|f_{\Lambda_m}\|^2$$

and

$$\begin{aligned} \|f_{\Lambda_m}\|^2 &= \|\Phi_m^{-1} [\langle R_{m-1}(f), \phi_i \rangle]_{i=1, \dots, n_m}^T\|^2 \\ &\geq \frac{1}{a_m^2} \|\langle R_{m-1}(f), \phi_i \rangle_{i=1, \dots, n_m}^T\|^2 \\ &\geq \frac{1}{a_m^2} r^2 n_m^2 \|R_{m-1}(f)\|^2. \end{aligned}$$

That is,

$$\|R_m(f)\|^2 = \|R_{m-1}(f)\|^2 - \|f_{\Lambda_m}\|^2 \leq \|R_{m-1}(f)\|^2 - \frac{1}{a_m^2} r^2 n_m^2 \|R_{m-1}(f)\|^2.$$

Summing the above inequality over  $m = 1, \dots, k$ , we get

$$\|R_k(f)\|^2 \leq \|R_0(f)\|^2 - r^2 \sum_{m=1}^k \frac{1}{a_m} n_m^2 \|R_{m-1}(f)\|^2 \leq \|f\|^2 - r^2 \sum_{m=1}^k \frac{1}{a_m} n_m^2 \|R_k(f)\|^2.$$

because of the monotonicity of  $\|R_m(f)\|$ . In other words,

$$(r^2 \sum_{m=1}^k \frac{1}{a_m} n_m^2 + 1) \|R_k(f)\|^2 \leq \|f\|^2.$$

As  $\sum_{m=1}^k n_m^2$  diverges and  $a_m$  nonincreases,  $\|R_k(f)\|$  has to converge to zero. This completes a proof of the convergence of this OGA. ■

The OGA can be used to solve our research problem Eq. (1). For  $\mathbf{y} \in \mathcal{R}_k$ , the OGA algorithm uses the indices which are associated with the terms  $|\langle \mathbf{y}, \phi_i \rangle|$ ,  $i \in \{1, 2, \dots, n\}$  which is  $\geq r\%$  of the largest value. As the size of  $\Lambda_i$  increases, it finds an approximation  $\mathbf{x}_{OGA, \epsilon}$  such that  $\Phi \mathbf{x}_{OGA, \epsilon}$  is closed to  $\mathbf{y}$  within the given  $\epsilon$ . That is,  $\|\Phi \mathbf{x}_{OGA, \epsilon} - \mathbf{y}\| \leq \epsilon$ .

We now explain why  $\mathbf{x}_{OGA, \epsilon}$  is a good approximation of  $\mathbf{x}$ . Due to the construction, the number of nonzero entries  $\|\mathbf{x}_{OGA, \epsilon}\|_0 = k^* \ll n$ . Similar to the RIP, let  $\alpha_k, \beta_k \geq 0$  be the best constants in the inequalities

$$\alpha_k \|\mathbf{z}\|_2 \leq \|\Phi \mathbf{z}\|_2 \leq \beta_k \|\mathbf{z}\|_2, \quad \text{for all } \|\mathbf{z}\|_0 \leq k.$$

Then since  $\|\mathbf{x} - \mathbf{x}_{OGA, \epsilon}\|_0 \leq k + k^*$ ,

$$\|\mathbf{x} - \mathbf{x}_{OGA, \epsilon}\|_2 \alpha_{k+k^*} \leq \|\Phi(\mathbf{x} - \mathbf{x}_{OGA, \epsilon})\|_2 = \|\mathbf{y} - \Phi \mathbf{x}_{OGA, \epsilon}\|_2 \leq \epsilon.$$

That is, we have

$$\|\mathbf{x} - \mathbf{x}_{OGA, \epsilon}\|_2 \leq \frac{\epsilon}{\alpha_{k+k^*}}.$$

In particular, when  $k^* \leq k$ , all we need is to assume that  $\alpha_{2k} > 0$  and hence  $\mathbf{x}_{OGA, \epsilon}$  is away from  $\mathbf{x}$  by  $\epsilon/\alpha_{2k}$ .

Next we need to show that  $k^* \leq k$  may happen. Assume that each column of  $\Phi$  is normalized. For  $\mathbf{y} = \Phi \mathbf{x}$  with  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , without loss of generality, we may assume that the support of  $\mathbf{x}$  is  $S = \{1, 2, \dots, k\}$ ,  $|x_k| = \min\{|x_j| \neq 0, j = 1, \dots, n\}$ , and  $|x_1| = \|\mathbf{x}\|_\infty$ .

Suppose that

$$k \leq \frac{1}{2M} + \frac{1}{2}, \quad (18)$$

where  $M = M(\Phi)$  stands for the mutual coherence of  $\Phi$ . Then we can claim that the support( $\mathbf{x}_{OGA, \epsilon}$ )  $\subset$  support( $\mathbf{x}$ ). Recall  $\Phi = [\phi_1, \dots, \phi_n]$  with  $\phi_i$

being the  $i$ th column of  $\Phi$ . Let us first compute the inner products of  $\mathbf{y}$  with  $\phi_i$ 's.

$$|\langle \mathbf{y}, \phi_i \rangle| = |\langle \Phi \mathbf{x}, \phi_i \rangle| = \left| \sum_{j=1}^k \langle x_j \phi_j, \phi_i \rangle \right|.$$

and

$$\left| \sum_{j=1}^k \langle x_j \phi_j, \phi_i \rangle \right| \geq |\langle x_1 \phi_1, \phi_i \rangle| - \sum_{j=2}^k |\langle x_j \phi_j, \phi_i \rangle|.$$

In particular, we have

$$|\langle \mathbf{y}, \phi_1 \rangle| \geq |x_1| - M(k-1)|x_2|.$$

and

$$|\langle \mathbf{y}, \phi_i \rangle| \leq \left| \sum_{j=1}^k \langle x_j \phi_j, \phi_i \rangle \right| \leq |x_1|kM.$$

By our assumption in Eq. (18), we have

$$|x_1| - M(k-1)|x_2| \geq |x_1| - M(k-1)|x_1| \geq |x_1|kM.$$

it follows that

$$|\langle \mathbf{y}, \phi_1 \rangle| \geq |\langle \mathbf{y}, \phi_i \rangle|, \quad \forall i \geq 2.$$

That is, the largest inner product is  $|\langle \mathbf{y}, \phi_1 \rangle|$ .

Furthermore, let us assume

$$k \leq \frac{1}{(1+r)M} \left( \frac{|x_k|}{|x_1|} + M \right) \quad \text{or} \quad rk|x_1|M \leq |x_k| - M(k-1)|x_1|, \quad (19)$$

where  $r$  is the positive constant  $r < 1$  employed in the OGA.

Then for  $2 \leq j \leq k$ ,  $|\langle \mathbf{y}, \phi_j \rangle| \geq |x_j| - M(k-1)|x_1| \geq |x_k| - M(k-1)|x_1|$  and  $r|\langle \mathbf{y}, \phi_1 \rangle| \leq rk|x_1|M$ . It follows that  $|\langle \mathbf{y}, \phi_j \rangle| \geq r|\langle \mathbf{y}, \phi_1 \rangle|$ . That is, the first greedy step in the above OGA picks up all the indices of the nonzero entries of  $\mathbf{x}$ .

In particular, when the nonzero entries of  $\mathbf{x}$  are 1 in absolute value, the condition in Eq. (19) is simplified to

$$k \leq \frac{1}{(1+r)M}(1+M).$$

That is, if  $k$  satisfies Eq. (18), then  $k$  satisfies Eq. (19). Under the condition in (18) or the conditions in (18) and (19), the OGA picks all the entries  $\phi_1, \dots, \phi_k$ . Hence, the  $\text{support}(\mathbf{x}_{OGA,\epsilon})$  is the same as the support of  $\mathbf{x}$ .

Furthermore,  $\|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|_1 \leq \sqrt{k}\|\mathbf{x}^* - \mathbf{x}\|_2$ . Since  $\|\Phi(\mathbf{x}_{OGA,\epsilon} - \mathbf{x})\|_2 = \|\Phi\mathbf{x}_{OGA,\epsilon} - \mathbf{y}\|_2 \leq \epsilon$ , we have

$$\begin{aligned} \epsilon^2 &\geq \|\Phi(\mathbf{x}_{OGA,\epsilon} - \mathbf{x})\|_2^2 \\ &= \|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|^2 + (\mathbf{x}_{OGA,\epsilon} - \mathbf{x})^T (G - I)(\mathbf{x}_{OGA,\epsilon} - \mathbf{x}) \\ &\geq \|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|_2^2 - M(\|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|_1^2 - \|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|_2^2) \\ &= (1 + M)\|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|_2^2 - Mk\|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|_2^2. \end{aligned}$$

That is,

$$\|\mathbf{x}_{OGA,\epsilon} - \mathbf{x}\|_2^2 \leq \frac{\epsilon^2}{1 - M(k - 1)}.$$

That is, under the assumption that the sparsity of  $\mathbf{x}$  is small, i.e., Eq. (18),  $\mathbf{x}_{OGA,\epsilon}$  approximates the sparse solution  $\mathbf{x}$  very well.

## 4.1 $L_1$ Greedy Algorithm

Recently, Kozlov and Petukhov proposed a new greedy algorithm (cf. [Kozlov and Petukhov'08]). It is called  $L_1$  Greedy Algorithm. The algorithm starts with the solution of the  $\ell_1$  minimization under the constraint  $A\mathbf{z}_0 = \mathbf{y}$ .

- [1] Let  $\mathbf{z}_0$  be the solution of the  $\ell_1$  minimization under the constraint  $A\mathbf{z} = \mathbf{y}$  among  $\mathbf{z} \in \mathbf{R}^n$ .
- [2] Let  $M = \|\mathbf{z}^0\|_\infty$ .
- [3] For  $i = 1, \dots, N$ , let  $W \in \mathbf{R}^n$  be a weighted vector with 1 in the all entries except for those entries which are  $1/10000$  when  $|\mathbf{z}_j^{i-1}| \geq 0.8M$ ,  $1 \leq j \leq n$ .
- [4] Solve the weighted  $\ell_1$  minimization problem

$$\min\left\{\sum_{j=1}^n |z_j|/w_j, A\mathbf{z} = \mathbf{y}, \mathbf{z} \in \mathbf{R}^n\right\}$$

and let  $\mathbf{z}^i$  be the solution.

[5] If  $\mathbf{z}^i$  is not yet a sparse solution, let  $M = 0.8M$  and return to Step 3.

The algorithm works well for random matrix  $A$  of size  $512 \times 1024$ . It is interesting to give an analysis of the convergence or reasons why the algorithm works.

## 5 The $\ell_q$ Approach

Let

$$\|\mathbf{x}\|_q = \left( \sum_{i=1}^n |x_i|^q \right)^{1/q}$$

be the standard  $\ell^q$  quasi-norm for  $0 < q < 1$ . It is easy to see that  $\lim_{q \rightarrow 0^+} \|\mathbf{x}\|_q^q = \|\mathbf{x}\|_0$ . We can use  $\|\mathbf{x}\|_q^q$  to approximate  $\|\mathbf{x}\|_0$ . Thus, we consider the following minimization

$$\min\{\|\mathbf{x}\|_q^q, \quad \mathbf{x} \in \mathbb{R}^n, \Phi\mathbf{x} = \mathbf{y}\}. \quad (20)$$

for  $0 < q \leq 1$  as an approximation of the original research problem Eq. (1). A solution of the above minimization is denoted by  $\Delta_q \Phi \mathbf{x}$ .

### 5.1 Recent Results on the $\ell_q$ Approach

The several  $\ell_q$  methods were studied recently in [Chartrand'07], [Foucart and Lai'08], [Davies and Gribonval'08] and [R. Saab and Ö. Yilmaz'08]. The first piece of results is shown in [Chartrand'07]

**Theorem 5.1** *Let  $q \in (0, 1]$ . Suppose that there exists a  $k > 1$  such that the matrix  $\Phi$  has RIP constant such that*

$$\delta_{ks} + k^{2/q-1} \delta_{(k+1)s} < k^{2/q-1} - 1.$$

*Then the solution of Eq. (20) is the sparsest solution.*

One can see that this result is a generalization of Theorem 3.10. When  $q = 1$  and  $k = 3$ , the above condition is the condition in Theorem 3.10. In fact, the proof is a generalization of the proof in [Candés, Romberg, and Tao'06] for  $\ell_1$  norm to  $\ell_q$  quasi-norm. In [Foucart and Lai'08], we felt that the non-homogeneity of the Restricted Isometry Property (13) contradicted the consistency of the problem with respect to measurement amplification,

or in other words, that it was in conflict with the equivalence of all the linear systems  $(cA)\mathbf{z} = c\mathbf{y}$ ,  $c \in \mathbb{R}$ . Instead, we introduce  $\alpha_k, \beta_k \geq 0$  to be the best constants in the inequalities

$$\alpha_k \|\mathbf{z}\|_2 \leq \|A\mathbf{z}\|_2 \leq \beta_k \|\mathbf{z}\|_2, \quad \|\mathbf{z}\|_0 \leq k.$$

Our results are to be stated in terms of a quantity invariant under the change  $A \leftarrow cA$ , namely

$$\gamma_{2s} := \frac{\beta_{2s}^2}{\alpha_{2s}^2} \geq 1.$$

In fact,  $\alpha_k^2 = 1 - \delta_k$  and  $\beta_k^2 = 1 + \delta_k$ . We use this slightly modified version of RIP and work through the arguments of [Candes, Romberg, Tao'06] in terms of quasi-norm  $\ell_q$  to get the following theorem.

Our main result in this section is the following (see [Foucart and Lai'08] for a proof)

**Theorem 5.2** *Given  $0 < q \leq 1$ , if*

$$\gamma_{2t} - 1 < 4(\sqrt{2} - 1) \left( \frac{t}{s} \right)^{1/q-1/2} \quad \text{for some integer } t \geq s, \quad (21)$$

*then every  $s$ -sparse vector is exactly recovered by solving Eq. (20).*

**Corollary 5.3** *Under the assumption that*

$$\gamma_{2s} < 4\sqrt{2} - 3 \approx 2.6569, \quad (22)$$

*every  $s$ -sparse vector is exactly recovered by solving (7).*

When  $q = 1$ , this result slightly improves Candès' condition in Theorem 3.9, since the constant  $\gamma_{2s}$  is expressed in terms of the Restricted Isometry Constant  $\delta_{2s}$  as

$$\gamma_{2s} = \frac{1 + \delta_{2s}}{1 - \delta_{2s}},$$

hence the condition (22) becomes  $\delta_{2s} < 2(3 - \sqrt{2})/7 \approx 0.4531$ .

The second special instance we are pointing out corresponds to the choice  $t = s + 1$ . In this case, Condition (21) reads

$$\gamma_{2s+2} < 1 + 4(\sqrt{2} - 1) \left( 1 + \frac{1}{s} \right)^{1/q-1/2}.$$

The right-hand side of this inequality tends to infinity as  $q$  approaches zero. The following result is then straightforward.



**Corollary 5.4** *Under the assumption that*

$$\gamma_{2s+2} < +\infty,$$

*every  $s$ -sparse vector is exactly recovered by solving (20) for some  $q > 0$  small enough.*

The key point is to show for any  $\mathbf{v}$  which is in the null space of  $\Phi$ , i.e.,  $\Phi\mathbf{v} = 0$ ,  $\|\mathbf{v}_S\|_q < \|\mathbf{v}_{\bar{S}}\|_q$  unless  $\mathbf{v} = 0$ , where  $S$  stands for the index set of the nonzero entries of the solution  $\mathbf{x} \in \mathcal{R}_k$ ,  $\mathbf{v}_S$  denotes the vector  $\mathbf{v}$  restricted in  $S$  with other entries being zero and  $\bar{S}$  is the complement indices of  $S$ .

This is indeed the case since for  $\mathbf{v} = \mathbf{x} - \mathbf{x}^*$  in the null space of  $\Phi$ ,  $\|\mathbf{v}_{\bar{S}}\|_q \leq \|\mathbf{v}_S\|_q$ , where  $\mathbf{x}^*$  is the solution of Eq. (20) and  $\mathbf{x}$  is the sparse vector supported on  $S$  satisfying  $\Phi\mathbf{x} = \mathbf{y}$ . Combining the above inequality, we have a contradiction that  $\|\mathbf{v}_S\|_q < \|\mathbf{v}_{\bar{S}}\|_q$  unless  $\mathbf{v} = 0$ . Thus, the solution of the minimization is the exact solution if we can show  $\|\mathbf{v}_S\|_q < \|\mathbf{v}_{\bar{S}}\|_q$ . This inequality was recognized in [Grinval and Nielson'03]. The condition (21) in Theorem 5.2 implies this inequality.

## 5.2 More about the $\ell_q$ Approach

We first show that the minimization problem Eq. (20) has a solution for  $q > 0$ . That is, the existence of the solution is independent of the RIP of  $\Phi$ . See [Foucart and Lai'08] for a proof.

**Theorem 5.5** *Fix  $0 < q < 1$ . There exists a solution  $\Delta_q \mathbf{A}\mathbf{x}$  solving Eq. (20).*

We next consider the situation that the measurements  $\mathbf{y}$  are imperfect. That is,  $\mathbf{y} = \Phi\mathbf{x}_0 + \mathbf{e}$  with unknown perturbation  $\mathbf{e}$  which is bounded by a known amount  $\|\mathbf{e}\|_2 \leq \theta$ . In this case we consider the following

$$\min\{\|\mathbf{x}\|_q^q, \quad \mathbf{x} \in \mathbb{R}^n, \|\Phi\mathbf{x} - \mathbf{y}\|_2 \leq \theta\}. \quad (23)$$

A solution of the above minimization is denoted by  $\Delta_{q,\theta}\Phi\mathbf{x}$ . As in the previous section, we have

**Theorem 5.6** *Fix  $0 < q < 1$  and  $\theta > 0$ . There exists a solution  $\Delta_{q,\theta}\Phi\mathbf{x}$  solving Eq. (23).*

In [Saab and Yilmaz'08], they extended the proof in [Candes'08] in the  $\ell_q$  setting. They have

**Theorem 5.7** *Let  $q \in (0, 1]$ . Suppose that  $\delta_{ks} + k^{2/q-1}\delta_{(k+1)s} < k^{2/q-1} - 1$  for some  $k > 1$  with  $ks \in \mathbf{Z}_+$ . Let  $\mathbf{x}^*$  be the solution of Eq. (23). Then*

$$\|\mathbf{x} - \mathbf{x}^*\|_2^q \leq C_1 \eta^p + \frac{C_2}{s^{1-q/2}} \Delta_s(\mathbf{x})_q^q$$

for two positive constants  $C_1$  and  $C_2$ .

Here, the quantity  $\Delta_k(\mathbf{x})_q$  denotes the error of best  $k$ -term approximation to  $\mathbf{x}$  with respect to the  $\ell_q$ -quasinorm, that is

$$\Delta_k(\mathbf{x})_q := \inf_{\|\mathbf{z}\|_0 \leq k} \|\mathbf{x} - \mathbf{z}\|_q.$$

The above theorem is an extension of Chartrand's result (cf. Theorem 5.1). Next we state another main theoretical result of this survey. We refer to [Foucart and Lai'08] for a proof.

**Theorem 5.8** *Given  $0 < q \leq 1$ , if Condition (21) holds, i.e. if*

$$\gamma_{2t} - 1 < 4(\sqrt{2} - 1) \left( \frac{t}{s} \right)^{1/q-1/2} \quad \text{for some integer } t \geq s, \quad (24)$$

then a solution  $\mathbf{x}^*$  of  $(P_{q,\theta})$  approximate the original vector  $\mathbf{x}$  with errors

$$\|\mathbf{x} - \mathbf{x}^*\|_q \leq C_1 \cdot \sigma_s(\mathbf{x})_q + D_1 \cdot s^{1/q-1/2} \cdot \theta, \quad (25)$$

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq C_2 \cdot \frac{\sigma_s(\mathbf{x})_q}{t^{1/q-1/2}} + D_2 \cdot \theta. \quad (26)$$

The constants  $C_1$ ,  $C_2$ ,  $D_1$ , and  $D_2$  depend only on  $q$ ,  $\gamma_{2t}$ , and the ratio  $s/t$ .

Comparison of the results in Theorems 5.8 and 5.7 is given in [Saab and Yilmaz'08]. It concludes that when  $k$  is around 2, the sufficient condition (24) is weaker while the condition in Theorem 5.7 is weaker when  $k > 2$ . Numerical experimental results in [Foucart and Lai'08] show that the  $\ell_q$  method is able to 100% recovery all the sparse vectors with sparsity is about  $s = m/2$ . That is, in order to have  $ks \leq m$ ,  $k$  is about 2.

Next we consider that negative results discussed in [Davies and Gribnoval'08]. That is, when the  $\ell_q$  method may fail.

**Theorem 5.9** *For any  $\epsilon > 0$ , there exists an integer  $s$  and dictionary  $\Phi$  with a restricted isometry constant  $\delta_{2s} \leq 1/\sqrt{2} + \epsilon$  for which  $\ell_1$  method fails on some  $k$  sparse vector.*

Now the gap between the positive result  $\delta_{2s} = 2(3 - \sqrt{2})/7 = 0.4531$  and the negative result  $\delta_{2s} = 1/\sqrt{2} + \epsilon = 0.7071$  is about 0.2540. In general, Davies and Gribnoval consider a special matrix  $\Phi$  which has a unit spectral norm, i.e.,

$$\|\Phi\|_2 = \sup_{\mathbf{y} \neq 0} \frac{\|\Phi \mathbf{y}\|_2}{\|\mathbf{y}\|_2} = 1.$$

Then they define

$$\sigma_k^2(\Phi) := \min_{\substack{\mathbf{y} \in \\ \|\mathbf{y}\|_0 \leq k}} \frac{\|\Phi \mathbf{y}\|_2}{\|\mathbf{y}\|_2}$$

which is equal to  $\alpha_k^2$  in [Foucart and Lai'08].

**Theorem 5.10** *Fix  $0 < q \leq 1$  and let  $0 < \eta_q < 1$  be the unique positive solution to*

$$\eta_q^{2/q} + 1 = \frac{2}{p}(1 - \eta_p).$$

*For any  $\epsilon > 0$ , there exist integers  $s \geq 1$ ,  $N \geq 2s + 1$  and a minimally redundant unit spectral norm tight frame  $\Phi_{N-1 \times N}$  with*

$$\sigma_{2s}^2(\Phi) \geq 1 - \frac{2}{2-q} \eta_q - \epsilon$$

*for which there exists an  $s$ -sparse vector which cannot be uniquely recovered by the  $\ell_q$  method.*

### 5.3 Numerical Computation of the $\ell_q$ Approach

The minimization problem  $(P_q)$  suggested to recover  $\mathbf{x}$  is nonconvex, Following [Foucart and Lai'08], we introduce an algorithm to compute a minimizer of the approximated problem, for which we give an informal but detailed justification.

We shall proceed iteratively, starting from a vector  $\mathbf{z}_0$  satisfying  $A\mathbf{z}_0 = \mathbf{y}$ , which is a reasonable guess for  $\mathbf{x}$ , and constructing a sequence  $(\mathbf{z}_n)$  recursively by defining  $\mathbf{z}_{n+1}$  as a solution of the minimization problem

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} \quad \sum_{i=1}^N \frac{|z_i|}{(|z_{n,i}| + \epsilon_n)^{1-q}} \quad \text{subject to} \quad A\mathbf{z} = \mathbf{y}. \quad (27)$$

Here, the sequence  $(\epsilon_n)$  is a nonincreasing sequence of positive numbers. It might be prescribed from the start or defined during the iterative process. In practice, we will take  $\lim_{n \rightarrow \infty} \epsilon_n = 0$ . We shall now concentrate on convergence issues. We start with the following

**Proposition 5.11** *For any nonincreasing sequence  $(\epsilon_n)$  of positive numbers and for any initial vector  $\mathbf{z}_0$  satisfying  $A\mathbf{z}_0 = \mathbf{y}$ , the sequence  $(\mathbf{z}_n)$  defined by (27) admits a convergent subsequence.*

Similar to the proof of Theorem 5.5, we can see that the solution of the above minimization exists. We further show that the solution  $\mathbf{x}_\epsilon$  of Eq.(27) will converge to the solution Eq. (20). For convenience, let  $\hat{\mathbf{x}}$  be a solution of Eq. (20).

**Theorem 5.12** *Fix  $0 < q \leq 1$ . Let  $\mathbf{x}_\epsilon$  be the solution of Eq. (27). Then  $\mathbf{x}_\epsilon$  converges to  $\hat{\mathbf{x}}$  as  $\epsilon \rightarrow 0_+$ .*

The new minimization problem Eq. (27) can be solved using  $\ell_1$  method since  $F_{q,\epsilon}(\mathbf{x})$  is a weighted  $\ell_1$  norm.

**Proposition 5.13** *Given  $0 < q < 1$  and the original  $s$ -sparse vector  $\mathbf{x}$ , there exists  $\eta > 0$  such that, if*

$$\epsilon_n < \eta \quad \text{and} \quad \|\mathbf{z}_n - \mathbf{x}\|_\infty < \eta \quad \text{for some } n, \quad (28)$$

*then the algorithm 27 produces the exact solution. That is,*

$$\mathbf{z}_k = \mathbf{x} \quad \text{for all } k > n.$$

*The constant  $\eta$  depends only on  $q$ ,  $\mathbf{x}$ , and  $\gamma_{2s}$ .*

**Lemma 5.14** *Given  $0 < q \leq 1$  and an  $s$ -sparse vector  $\mathbf{x}$ , if Condition (21) holds, i.e. if*

$$\gamma_{2t} - 1 < 4(\sqrt{2} - 1) \left( \frac{t}{s} \right)^{1/q-1/2} \quad \text{for some integer } t \geq s,$$

*then for any vector  $\mathbf{z}$  satisfying  $A\mathbf{z} = \mathbf{y}$ , one has*

$$\|\mathbf{z} - \mathbf{x}\|_q^q \leq C [\|\mathbf{z}\|_q^q - \|\mathbf{x}\|_q^q],$$

*for some constant  $C$  depending only on  $q$ ,  $\gamma_{2t}$ , and the ratio  $s/t$ .*

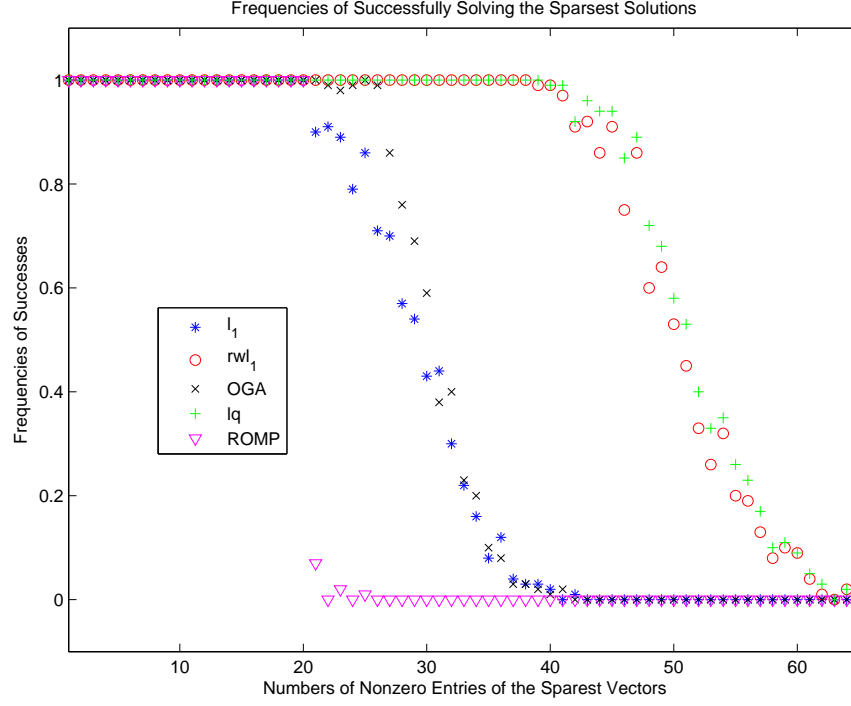


Figure 1: Comparison of  $\ell_1$ ,  $\ell_q$ , and OGA methods for sparsest solutions

Finally we present the following

**Proposition 5.15** *Given  $0 < q < 1$  and the original  $s$ -sparse vector  $\mathbf{x}$ , if Condition (21) holds, i.e. if*

$$\gamma_{2t} - 1 < 4(\sqrt{2} - 1) \left( \frac{t}{s} \right)^{1/q - 1/2} \quad \text{for some integer } t \geq s,$$

*then there exists  $\zeta > 0$  such that, for any nonnegative  $\epsilon$  less than  $\zeta$ , the vector  $\mathbf{x}$  is exactly recovered by solving (27). The constant  $\zeta$  depends only on  $N$ ,  $q$ ,  $\mathbf{x}$ ,  $\gamma_{2t}$ , and the ratio  $s/t$ .*

Numerical results show that our  $\ell_q$  approximation method works well. In Figure 1, we present the frequencies of the exact recovery using various methods for Gaussian random matrix of size  $128 \times 512$  for various sparse vectors. For each sparsity, we randomly generate the Gaussian random matrix  $\Phi$  and

a vector  $\mathbf{x}$  with the given sparsity and tested various methods to solve the  $\mathbf{x}$  for 100 times. The number of exact recovery by each method is divided by 100 to obtain the frequency for the method.

## References

- [1] Baraniuk, R., M. Davenport, R. DeVore, and M. B. Wakin, A simple proof of the restricted isometry property for random matrices, *Constructive Approximation*, to appear, 2008.
- [2] Buldygin, V. V. and Yu, V. Kozachenko, *Metric Characterization of Random Variables and Random Processes*, AMS Publication, Providence, 2000.
- [3] Candés, E. J. Compressive sampling. *International Congress of Mathematicians. Vol. III*, 1433–1452, Eur. Math. Soc., Zürich, 2006.
- [4] Candés, E. J., J. K. Romberg, Quantitative robust uncertainty principles and optimally sparse decompositions. *Found. Comput. Math.* 6 (2006), 2, 227–254.
- [5] Candés, E. J., J. K. Romberg, J. K. and T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.* 59 (2006), 1207–1223.
- [6] Candés, E. J. and T. Tao, Decoding by linear programming, *IEEE Trans. Inform. Theory* 51 (2005), no. 12, 4203–4215.
- [7] Candés, E. J. and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies, *IEEE Trans. Inform. Theory* 52 (2006), no. 12, 5406–5425.
- [8] Candés, E. J., M. Watkin, and S. Boyd, Enhancing Sparsity by Reweighted  $l_1$  Minimization, manuscript, 2007.
- [9] R. Chartrand, Nonconvex compressed sensing and error correction, in *International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 2007.

- [10] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, *IEEE Signal Process. Letters*, 14(2007), 707–710.
- [11] R. Chartrand and V. Staneva, Restricted isometry properties and non-convex compressive sensing, *Inverse Problem*, to appear, 2008.
- [12] M. Davies and Rémi Gribonval, Restricted Isometry constants where  $\ell_q$  sparse recovery can fail for  $0 < q \leq 1$ , manuscript, 2008.
- [13] Donoho, D. L., Compressed sensing, *IEEE Trans. Inform. Theory* 52 (2006), 1289–1306.
- [14] Donoho, D. L., Sparse components of images and optimal atomic decompositions. *Constr. Approx.* 17 (2001), 353–382.
- [15] D. L. Donoho, Unconditional bases are optimal bases for data compression and for statistical estimation, *Appl. Comput. Harmonic Anal.*, 1(1993), 100–115.
- [16] D. L. Donoho, Unconditional bases and bit-level compression, *Appl. Comput. Harmonic Anal.*, 3(1996), pp. 388–92.
- [17] Donoho, D. L. and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via  $l^1$  minimization, *Proc. Natl. Acad. Sci. USA* 100 (2003), no. 5, 2197–2202.
- [18] Donoho, D. L., M. Elad, and V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 52 (2006), 6–18.
- [19] Donoho, D. L. and J. Tanner, Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proc. Natl. Acad. Sci. USA* 102 (2005), no. 27, 9446–9451.
- [20] Elad, M. and A. M. Bruckstein, *IEEE Trans. Inf. Theory* 48(2002), 2558–2567.
- [21] S. Foucart and M. J. Lai, Sparsest Solutions of Underdetermined Linear Systems via  $\ell_q$  minimization for  $0 < q \leq 1$ , to appear in *Applied Comput. Harmonic Analysis*, 2009.

- [22] Geman, S., A limit theorem for the norm of random matrices, *Ann. Prob.*, 8(1980), 252–261.
- [23] Gribnoval, R and M. Nielsen, Sparse decompositions in unions of bases, *IEEE Trans. Info. Theory*, 49(2003), 3320–3325.
- [24] I. Kozlov and A. Petukhov,  $\ell_1$  greedy algorithm for sparse solutions of underdetermined linear system, manuscript, 2008.
- [25] M. J. Lai, Restricted isometry property for sub-Gaussian random matrices, manuscript, 2008.
- [26] Ledoux, M., *The concentration of measure phenomenon*, AMS Publication, Providence, 2001.
- [27] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and subgaussian ensembles, *Constructive Approx.* 28(2008) 277–289.
- [28] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Reconstruction and subgaussian operators in asymptotic geometric analysis, *Geometric and Functional Analysis* 17(2007), 1248–1272.
- [29] Natarajan, B. K., Sparse approximate solutions to linear systems, *SIAM J. Comput.*, vol. 24, pp. 227234, 1995.
- [30] Needell, D. and R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, manuscript, 2007.
- [31] Petukhov, A., Fast implementation of orthogonal greedy algorithm for tight wavelet frames, *Signal Processing*, 86(2006), 471–479.
- [32] G. Pisier, *Probabilistic Methods in the Geometry of Banach Spaces*, Springer Verlag, Lecture Notes in Mathematics, No. 1206, 1986.
- [33] R. Saab and Ö. Yilmaz, Sparse recovery by nonconvex optimization – instant optimality, manuscript, 2008.
- [34] Temlyakov, V. N., Weak greedy algorithms, *Adv. Comput. Math.* 12 (2000), 213–227.



- [35] Temlyakov, V. N., Nonlinear methods of approximation, Foundations of Comp. Math., 3 (2003), 33–107.
- [36] Tropp, J. A., Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inf. Theory, 50 (2004), 2231–2242.
- [37] Wachter, K. W., The strong limits of random matrix spectra for sample matrices of independent elements, Ann. Prob., 6(1978), 1–18.

## 6 Appendix 1: Gaussian Random Matrices

Let  $A = [a_{ij}]_{1 \leq i \leq m, 1 \leq j \leq n}$  be a rectangular matrix with  $a_{ij}$  being iid Gaussian random variables with mean zero and variance  $\sigma^2$ . Let  $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbf{R}^n$  be a vector. We use  $\|\mathbf{x}\|_2$  denotes the norm of  $\mathbf{x}$ . Consider a random variable  $X = (X_1, \dots, X_m)^T$  with  $X_i = (\sum_{j=1}^n a_{ij}x_j)^2, i = 1, \dots, m$ . Since  $\mathbf{E}(a_{ij}) = 0$ , we have  $\mathbf{E}(X_i) = \sigma^2 \|\mathbf{x}\|_2^2$  for all  $i$ . Let  $\xi_i = X_i - \mathbf{E}(X_i)$  be a new random variable and let

$$S_m = \sum_{i=1}^m \xi_i$$

be the sum of these new independent random variables. It is easy to see that

$$S_m = \|A\mathbf{x}\|_2^2 - m\sigma^2 \|\mathbf{x}\|_2^2.$$

In this section, we are interested in proving the following inequality.

**Theorem 6.1** *For any  $\epsilon > 0$ , the probability*

$$\mathcal{P}(|\|A\mathbf{x}\|_2^2 - m\sigma^2 \|\mathbf{x}\|_2^2| < \epsilon \|\mathbf{x}\|_2^2) \geq 1 - 2 \exp\left(-\frac{\epsilon^2 m}{(m\sigma^2)(c\epsilon + 2m\sigma^2)}\right), \quad (29)$$

where  $c$  is a positive constant independent of  $\epsilon$  and  $\|\mathbf{x}\|_2$ .

We plan to use the Bernstein inequality (cf. [Buldygin and Kozachenko'00, p.27]) to prove this result. For convenience, we state the inequality below.

**Theorem 6.2** *Suppose that  $\xi_i, 1 \leq i \leq m$  are independent random variables with  $\mathbf{E}(\xi_i) = 0$  and  $\mathbf{E}(\xi_i^2) = \nu_i^2 < \infty, 1 \leq i \leq m$ . Let  $S_m = \sum_{i=1}^m \xi_i$ . Moreover, suppose that there exists a constant  $H > 0$  such that*

$$|\mathbf{E}(\xi_i^k)| \leq \frac{m!}{2} \nu_i^2 H^{k-2} \quad (30)$$

for all integer  $k > 1$  and all  $i = 1, \dots, m$ . Then the following inequality holds for all  $t > 0$ : the probability

$$\mathcal{P}(|S_m| > t) \leq \exp \left\{ -\frac{t^2}{2(tH + \sum_{i=1}^m \nu_i^2)} \right\}.$$

**Proof. (The proof of Theorem 6.1.)** We need to study Eq. (30) for  $\xi_i = X_i - \mathbf{E}(X_i)$  for  $k \geq 3$  since for  $k = 2$ , Eq. (30) is satisfied trivially.

For convenience, let  $\mu = \mathbf{E}(X_i) = \sigma^2 \|\mathbf{x}\|_2^2$ . It is easy to see  $\mathbf{E}(|\xi_i|^2) = 2\mu^2$ . Thus,  $\nu_i^2 = 2\mu^2$ . For  $k \geq 3$ , we have

$$\mathbf{E}(|\xi_i|^k) = \mathbf{E}((X_i - \mu)^k) = \sum_{j=0}^k \binom{k}{j} \mathbf{E}(X_i^j) (-1)^{k-j} \mu^{k-j}.$$

Let us spend some effort to compute  $\mathbf{E}(X_i^j)$ . We have

$$\mathbf{E}(X_i^j) = \mathbf{E}\left(\sum_{j=1}^n a_{ij} x_j\right)^{2j} = \sum_{j_1 + \dots + j_n = 2j} \frac{(2j)!}{j_1! \dots j_n!} \mathbf{E}(a_{i,1}^{j_1} a_{i,2}^{j_2} \dots a_{i,n}^{j_n}) x_1^{j_1} \dots x_n^{j_n}.$$

Note that  $\mathbf{E}(a_{ij}^\ell) = 0$  for all odd integers  $\ell$  and it is known (using integration by parts) that  $\mathbf{E}(a_{ij}^\ell) = \frac{\ell!}{2^{\ell/2}(\ell/2)!} \sigma^\ell$  for even integers  $\ell$ . Since  $a_{ij}$  are iid random variables, we have

$$\begin{aligned} \mathbf{E}(X_i^j) &= \sum_{2j_1 + \dots + 2j_n = 2j} \frac{(2j)!}{(2j_1)! \dots (2j_n)!} \mathbf{E}(a_{i,1}^{2j_1} a_{i,2}^{2j_2} \dots a_{i,n}^{2j_n}) (x_1)^{2j_1} \dots (x_n)^{2j_n} \\ &= \frac{(2j)!}{j!} \sum_{j_1 + \dots + j_n = j} \frac{j!}{(2j_1)! \dots (2j_n)!} \frac{(2j_1)! \dots (2j_n)!}{2^j j_1! \dots j_n!} \sigma^{2j} (x_1)^{2j_1} \dots x_n^{2j_n} \\ &= \frac{(2j)! \sigma^{2j}}{2^j j!} \left( \sum_{j=1}^n x_j^2 \right)^j \\ &= \frac{(2j)!}{2^j j!} \sigma^{2j} \|\mathbf{x}\|_2^{2j} = \frac{(2j)!}{2^j j!} \mu^j. \end{aligned}$$

Thus,

$$\mathbf{E}(|\xi_i|^k) \leq \sum_{j=0}^k \binom{k}{j} \frac{(2j)!}{2^j j!} \mu^j \mu^{k-j}.$$

By using Stirling's formula, we have  $\frac{(2j)!}{2^j j!} \leq 2^j j!/2 \leq 2^j k!/2$  and hence,

$$\begin{aligned} |\mathbf{E}(|\xi_i|^k)| &\leq \sum_{j=0}^k \binom{k}{j} \frac{2^j k!}{2} \mu^k \\ &= \frac{k!}{2} 3^k \mu^k = \frac{k!}{2} 2\sigma^4 \|\mathbf{x}\|_2^4 \frac{9}{2} 3^{k-2} (\sigma^2 \|\mathbf{x}\|_2^2)^{k-2} \\ &\leq \frac{k!}{2} 2\sigma^4 \|\mathbf{x}\|_2^4 H^{k-2} \end{aligned}$$

with  $H = 13.5\sigma^2 \|\mathbf{x}\|_2^2$ . That is, Eq. (30) is satisfied for  $k \geq 3$ . By Theorem 6.2, we have

$$P(|\|\mathbf{A}\mathbf{x}\|_2^2 - m\sigma^2 \|\mathbf{x}\|_2^2| > t) \leq 2 \exp \left\{ -\frac{t^2}{2(t13.5\mu + 2m\mu^2)} \right\}. \quad (31)$$

Choosing  $t = \epsilon \|\mathbf{x}\|_2^2$ , we have  $t13.5\mu + 2m\mu^2 = \sigma^2 \|\mathbf{x}\|_2^4 (13.5\epsilon + 2m\sigma^2)$  and the above probability yields:

$$P(|\|\mathbf{A}\mathbf{x}\|_2^2 - m\sigma^2 \|\mathbf{x}\|_2^2| > \epsilon \|\mathbf{x}\|_2^2) \leq 2 \exp \left\{ -\frac{\epsilon^2 m}{2(m\sigma^2)(13.5\epsilon + 2m\sigma^2)} \right\}. \quad (32)$$

In other word, the desirable result of Theorem 6.1 is proved. ■

We remark that when  $\sigma = 1/\sqrt{m}$ , the estimate Eq. (29) gives a proof of Theorem 3.11. For this special case, we have

**Theorem 6.3** *Suppose that  $\xi$  is a Gaussian random variable with mean zero and variance  $\sigma^2$ . Let  $A$  be an  $m \times n$  matrix whose entries are iid copies of  $\xi$ . For any  $\epsilon > 0$ , the probability*

$$\mathcal{P}(|\|\mathbf{A}\mathbf{x}\|_2^2 - m\sigma^2 \|\mathbf{x}\|_2^2| < \epsilon n\sigma^2 \|\mathbf{x}\|_2^2) \geq 1 - 2 \exp \left\{ -\frac{\epsilon^2 m}{c} \right\}, \quad (33)$$

where  $c$  is a positive constant independent of  $\epsilon$  and  $\|\mathbf{x}\|_2$ .

**Proof.** (The Proof of Theorem 6.3.) For a random matrix  $A$  of size  $m \times n$  with entries  $a_{ij}$  being iid Gaussian random variables with zero mean and variance  $\sigma^2$ , then we use  $\epsilon m\sigma^2$  for  $\epsilon$  in Eq. (32). Then we have

$$\mathcal{P}(|\|\mathbf{A}\mathbf{x}\|_2^2 - m\sigma^2 \|\mathbf{x}\|_2^2| > \epsilon m\sigma^2 \|\mathbf{x}\|_2^2) \leq 2 \exp \left\{ -\frac{\epsilon^2 m}{2(\epsilon 13.5 + 2)} \right\}. \quad (34)$$

This completes a proof of Theorem 6.3. ■

## PATTERNS $P$

$$\emptyset \subseteq P \subseteq (\mathbb{Q} \times \mathbb{Q}) \cap ([0,1] \times [0,1])$$

Milton del Castillo Lesmes Acosta

Universidad Distrital Francisco José de Caldas, Bogotá, Colombia.

**ABSTRACT.** We will find, from an approximation point of view, patterns, from uniform patterns to fractal patterns.

### 1. RATIONAL NUMBERS.

Given the set  $\mathbb{Q} = \left\{ \frac{p}{q} \mid p \in \mathbb{Z}, q \in \mathbb{Z}, q \neq 0 \right\}$ ,  $\mathbb{Z}$  the set of integers. We will consider

$$\mathbb{Q}_n = \left\{ \frac{m}{n} \mid m = 0, 1, 2, \dots, n \right\} \text{ for } n = 1, 2, \dots \text{ then}$$

$$\mathbb{Q} \cap [0,1] = \bigcup_{n=1}^{\infty} \mathbb{Q}_n$$

In one-dimensional finite elements a 10-element model of  $[0,1]$  is illustrated in Fig. 1 where the connectivity is important to the global model.

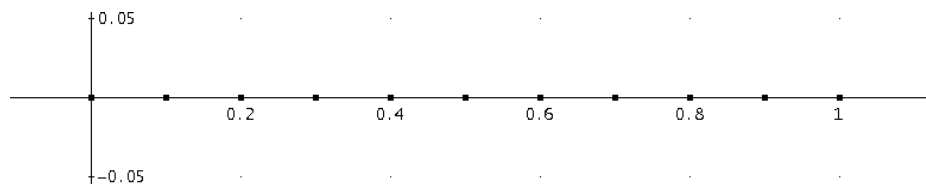


Fig. 1

In two-dimensional finite elements a model of  $[0,1] \times [0,1]$  is illustrated in Fig. 2, connectivity becomes more complicated (triangular elements is more popular)

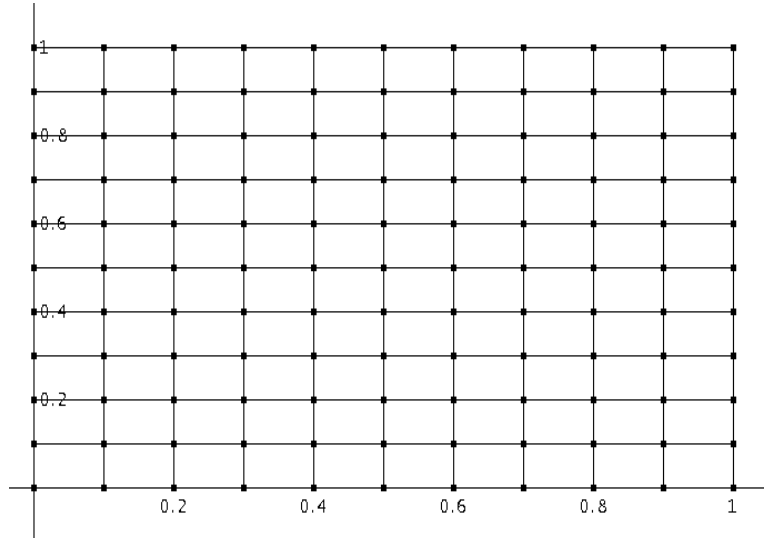


Fig. 2

The sets  $\mathbb{Q}$  and  $\mathbb{Q} \times \mathbb{Q}$  (and so on) appear similar to all levels of magnification. Then we will follow a construction, specifically of  $\mathbb{Q} \cap [0,1] = \bigcup_{n=1}^{\infty} \mathbb{Q}_n$  as a projection of the set  $\bigcup_{n=1}^{\infty} \mathbb{Q}_n \times \mathbb{Q}_n$ . The set  $\mathbb{Q}_{10} \times \mathbb{Q}_{10}$  is as illustrated in Fig. 3

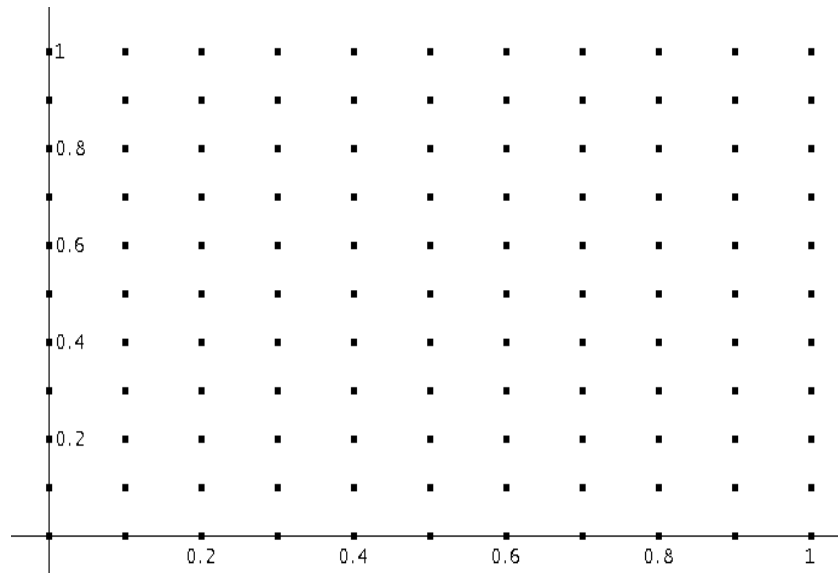


Fig. 3

And in the following figures, Fig. 4, Fig. 5, Fig. 6, there are cases for different values of  $p$  in

$$\bigcup_{n=1}^p \mathbb{Q}_n \times \mathbb{Q}_n$$

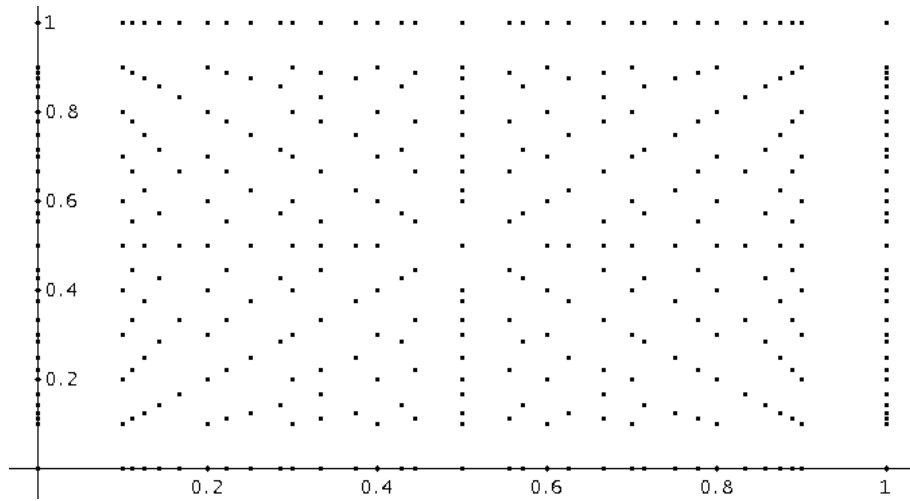


Fig. 4

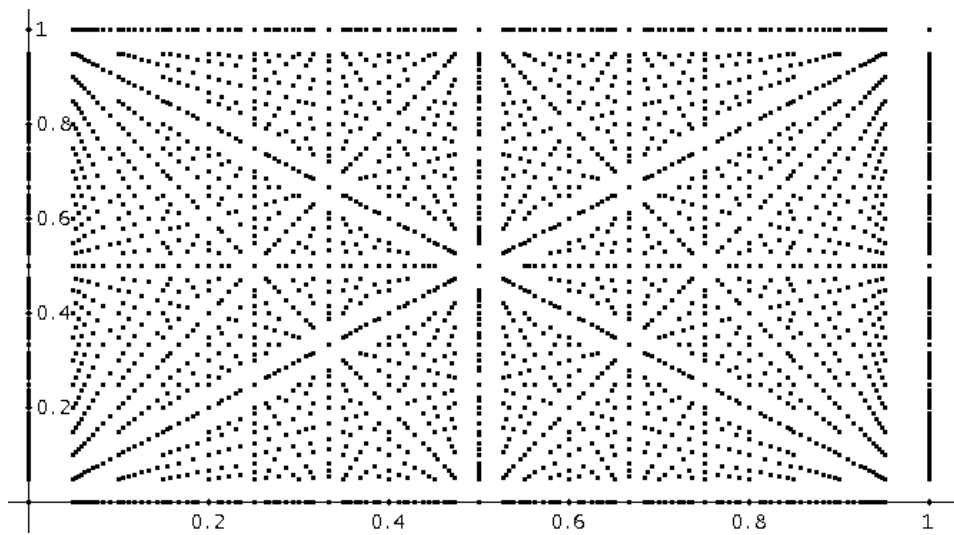


Fig. 5

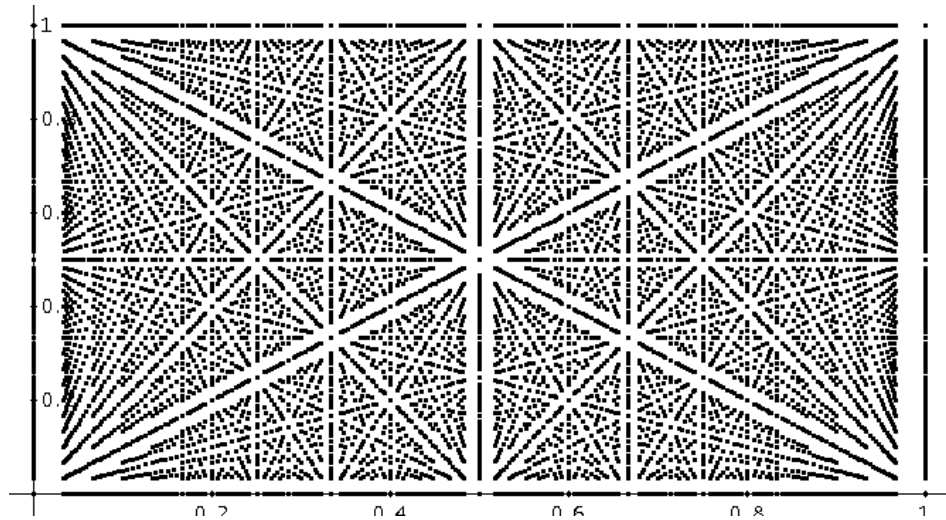


Fig. 6

This sets are specifically the sets of points  $\left\{ \left( \frac{a}{c}, \frac{b}{c} \right) \left| \begin{array}{l} a = 0, 1, \dots, c \\ b = 0, 1, \dots, c \\ c = 1, 2, \dots, p \end{array} \right. \right\}$ , for a specified  $p \in \mathbb{Z}^+$ .

Patterns due to transformations defined for functions  $f$  and  $g$  of the type  $(x, y) \rightarrow (f(x, y), g(x, y))$  over a pattern, carry new patterns as in the Fig. 7, Fig. 8, Fig. 9, Fig.

10, applying for example  $(x, y) \rightarrow (x^2, y^2)$  over the set  $\bigcup_{n=1}^p \mathbb{Q}_n \times \mathbb{Q}_n$

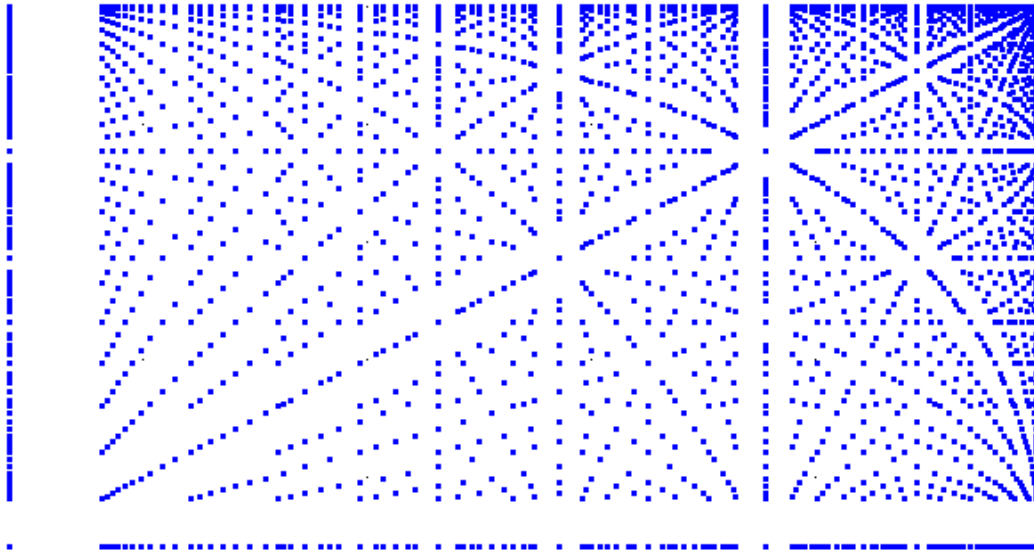


Fig. 7

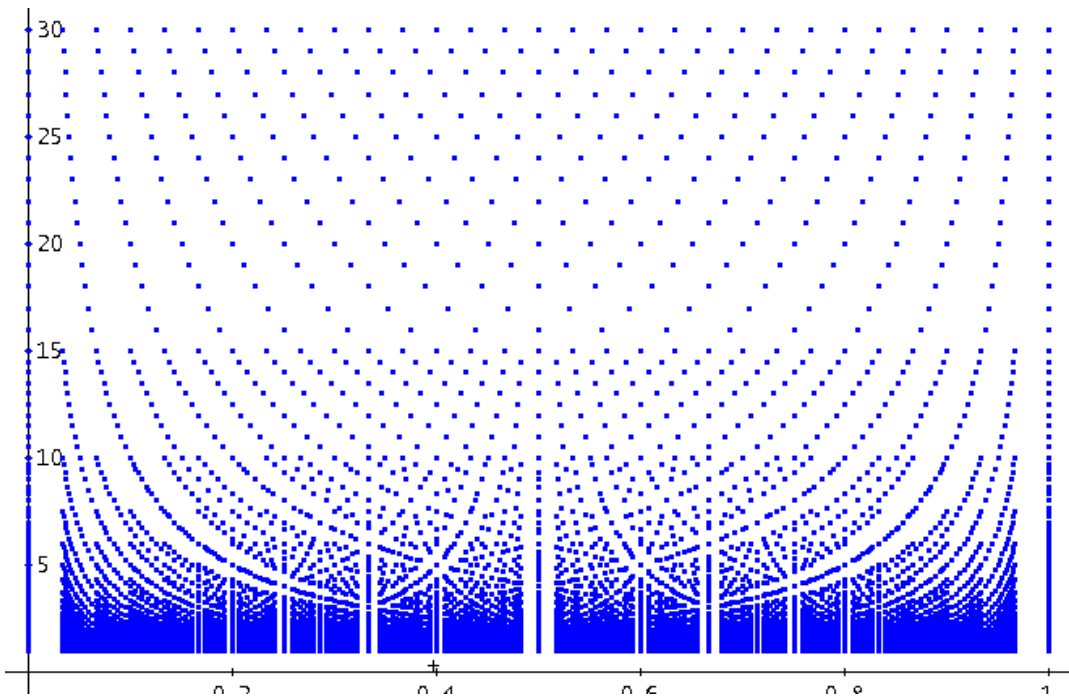


Fig. 8

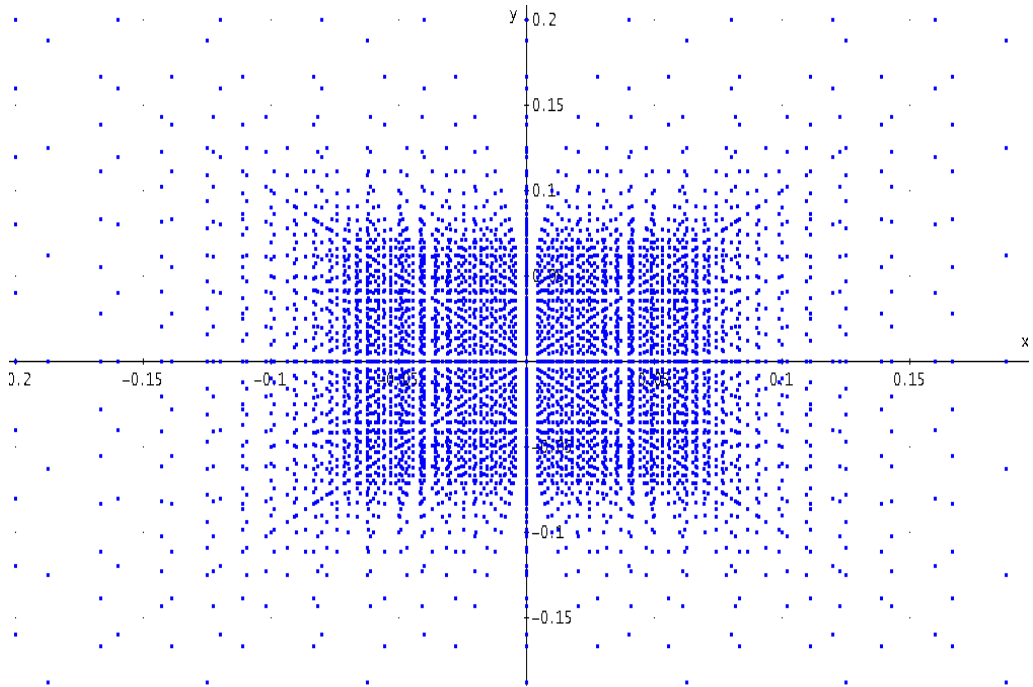


Fig. 9



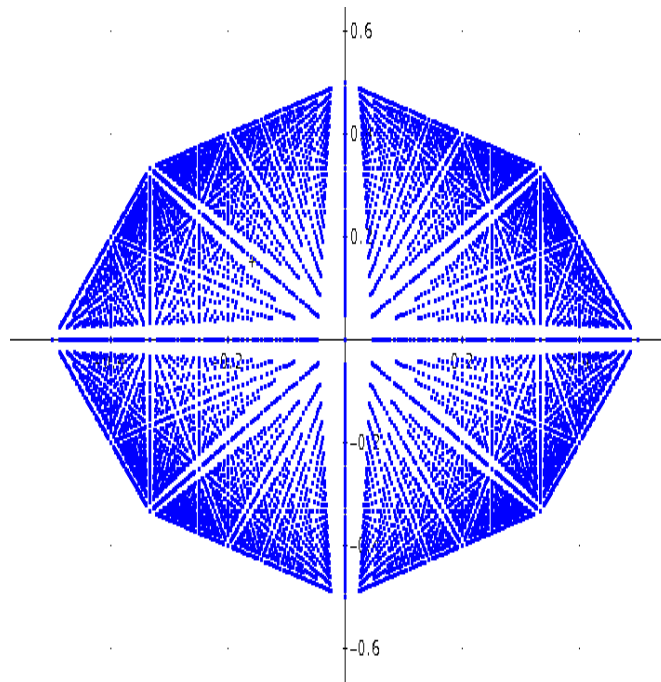


Fig. 10

Working with other sets for example Farey sequences, we can get more patterns like the following in Fig. 11



Fig. 11

## 2. NEW PATTERNS FROM PATTERNS.

To construct a periodic function  $f : \mathbb{R} \rightarrow \mathbb{R}$  with the condition  $f(x+1) = f(x)$  we need only to have  $g : \mathbb{R} \rightarrow \mathbb{R}$  with  $g(x) = f(x)$  for  $x \in [0,1)$  and then  $f(x) = g(\bar{x} \equiv x \pmod{1})$ , where  $\bar{x} \equiv x \pmod{1}$  means that for  $k \in \mathbb{Z}$  we have  $\bar{x} = x - k \in [0,1)$ . In another way the same thing is obtained if we write  $f(x) = g(x - \llbracket x \rrbracket)$  using the integer part function.

## 3. REPEATING A LOCATED PATTERN TO FIND NEW PATTERNS.

Take the family of lines from the point  $[0,1]$  to  $\left[\frac{1}{2}, a\right]$  and from the point  $[1,1]$  to  $\left[\frac{1}{2}, a\right]$  defined as

$$\det \begin{bmatrix} x & y & 1 \\ 0 & 1 & 1 \\ \frac{1}{2} & a & 1 \end{bmatrix} = 0, \quad \det \begin{bmatrix} x & y & 1 \\ 1 & 1 & 1 \\ \frac{1}{2} & a & 1 \end{bmatrix} = 0$$

that is  $y = 2x(a-1)+1$  and  $y = 2x(1-a)+2a-1$  respectively for  $a = 0, 0.1, 0.2, \dots, 1$  with the restriction  $x \in [0,1)$  as is shown in Fig. 12.

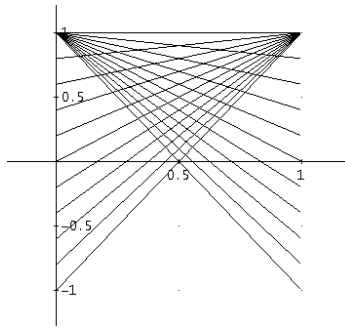


Fig. 12

then  $y - \llbracket y \rrbracket = 2(x - \llbracket x \rrbracket)(a-1)+1$  and  $y - \llbracket y \rrbracket = 2(x - \llbracket x \rrbracket)(1-a)+2a-1$  for  $a = 0, 0.1, 0.2, \dots, 1$  produce the pattern as shown in Fig. 13

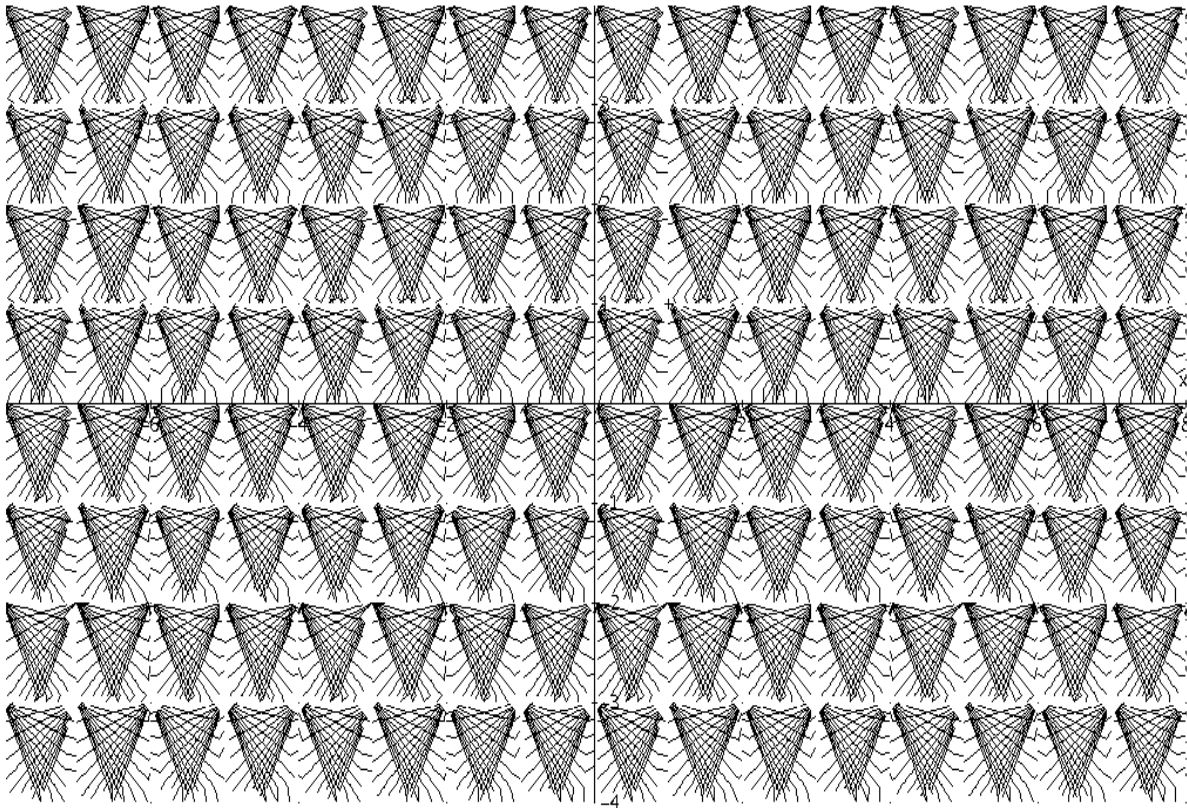


Fig. 13

#### 4. BIBLIOGRAPHY

Barnsley, Michael F. , and Hawley Rising. *Fractals Everywhere*. Boston: Academic Press Professional, 1993.

Fugal D. Lee. *Conceptual Waveletes in digital signal processing*. 2007

*Mandelbrot, Benoît B. The Fractal Geometry of Nature*. New York: W. H. Freeman and Co., 1982

Peitgen, Heinz-Otto, and Dietmar Saupe, eds. *The Science of Fractal Images*. New York: Springer-Verlag, 1988

The Derive - Newsletter. Derive Users Group. 71/72. Austria 2008

# A note on the construction of the $sl(2, R)$ integral for ordinary differential equations of maximal symmetry

Sibusiso Moyo

Department of Mathematics, Statistics and Physics,  
Steve Biko Campus, Durban University of Technology,  
PO Box 953, Durban 4000, South Africa.  
e-mail:moyos@dut.ac.za

March 15, 2009

We discuss some of the integrability properties of ordinary differential equations of maximal symmetry. The relationship between the solution symmetries, the fundamental integrals and the  $sl(2, R)$  integrals is discussed. A linear combination of the fundamental integrals also leads to an integral which has the  $sl(2, R)$  subalgebra. The construction of the  $sl(2, R)$  integral for order  $n \geq 9$  can be tedious and hence requires a more compact generating function. Here an illustration of such a construction is given and the integral is constructed from the direct integration of the differential equation after multiplying by one of the autonomous integrating factors. It is also important to note that symmetries and first integrals form the underlying mathematical basis for the algebraic theory of integrable equations.

Keywords: Symmetry, Lie-algebra, maximal symmetry, Integral

## 1 Introduction

We recall that Lie ([1], p 405) showed that the maximum number of point symmetries for second order ordinary differential equations is  $2 + 6$  and for higher order equations ( $n \geq 3$ ) is  $n + 4$ . In addition he further showed that an  $n$ th order equation which possesses  $n + 4$  point symmetries is equivalent to  $y^{(n)} = 0$ , where  $^{(n)}$  denotes  $d^n/dx^n$ , under a point transformation

$$X = F(x, y) \quad Y = G(x, y).$$

For  $n \geq 3$ ,  $X = F(x)$  which is called a fibre preserving transformation [2]. Lie showed that every linearisable second-order ordinary differential equation has the form

$$y'' = p(x, y)y'^3 + q(x, y)y'^2 + r(x, y)y' + s(x, y), \quad (1)$$

where the coefficients  $p, q, r$  and  $s$  satisfy the conditions  $A = 0$  and  $B = 0$ , where

$$\begin{aligned} A &= 2q_{xy} - 3p_{xx} - r_{yy} - 3p_x r + 3p_y s + 2q_x q - 3r_x p - r_y q + 6s_y p \\ B &= 2r_{xy} - q_{xx} - 3s_{yy} - 6p_x s + q_y s + 3q_y s - 2r_y r - 3s_x p + 3s_y q. \end{aligned}$$

It is well known that when a symmetry is used to determine a first integral for a differential equation, the symmetry provides an integrating factor for the equation and remains as a symmetry of the first integral. A considerable amount of work on constructing integrating factors for ordinary differential equations has been done including the works by Cheb-Terrab and Roche in 1999 [4] and Leach and Bouquet in 2002 [6]. From these works one concludes that there is a strong link between the existence of symmetries be they local or nonlocal and the integrability of the given equation or systems of equations.

## 2 First Integrals and Integrating Factors

For purposes of this discussion we give a definition of a symmetry of a differential equation as follows:

**Definition:** An  $n$ th order ordinary differential equation

$$E(x, y, y', \dots, y^{(n)}) = 0, \quad (2)$$

admits the one-parameter Lie-group of point transformations

$$\begin{aligned} \bar{x} &= X(x, y; \epsilon) = x + \epsilon \xi(x, y) + O(\epsilon^2) \\ \bar{y} &= Y(x, y; \epsilon) = y + \epsilon \eta(x, y) + O(\epsilon^2), \end{aligned} \quad (3)$$

with infinitesimal generator

$$G = \xi(x, y)\partial_x + \eta(x, y)\partial_y \quad (4)$$

if the condition

$$G^{[n]} \left[ E(x, y, y', \dots, y^{(n)}) \right]_{E=0} = 0 \quad (5)$$

holds, where  $G^{[n]}$  is the  $n$ th extension of  $G$  needed to transform the derivatives in  $E = 0$  given by

$$G^{[n]} = G + \sum_{i=1}^n \left\{ \eta^{(i)} - \sum_{j=0}^{i-1} \binom{i}{j} y^{(j+1)} \xi^{(i-j)} \right\} \partial_{y^{(i)}}. \quad (6)$$

Here the indices denote total differentiation *with respect to*  $x$ . (See Bluman and Kumei [3].)

**Definition:** In addition a first integral  $I$  for an equation of maximal symmetry  $E = y^{(n)} = 0$  is defined as

$$I = f(y, y', y'', \dots, y^{(n-1)})$$

where

$$\frac{dI}{dx}|_{E=0} = 0 \iff \frac{df}{dx}|_{E=0} = 0. \quad (7)$$

This means that if  $h(x, y, y', y'', \dots, y^{(n-1)})$  is an integrating factor then

$$\frac{dI}{dx}|_{E=0} = hE(x, y, y', \dots, y^{(n)})|_{E=0} = 0. \quad (8)$$

### 3 Some algebraic properties of equations of maximal symmetry

According to Lie's classification [1] the real unimodular group  $sl(2, R)$  is provided by the Lie algebras spanned by the vectors

$$\partial_x, \quad x\partial_x + y\partial_y, \quad x^2\partial_x + 2xy\partial_y. \quad (9)$$

or

$$\partial_x, \quad x\partial_x + y\partial_y, \quad (x^2 - y^2)\partial_x + 2xy\partial_y. \quad (10)$$

As an example, we consider the well-known third order ordinary differential equation of maximal symmetry

$$y''' = 0 \quad (11)$$

which has seven Lie-point symmetries given as

$$\begin{aligned} G_1 &= \partial_y, & G_2 &= x\partial_y, & G_3 &= x^2\partial_y \\ G_4 &= y\partial_y \\ G_5 &= \partial_x, & G_6 &= x\partial_x + y\partial_y, & G_7 &= x^2\partial_x + 2xy\partial_y \end{aligned}$$

with algebra  $3A_1, \{sl(2, R) \oplus_s A_1\}$ .

**Remark** The first three are solution symmetries and provide a basis for the solution of (11), that is,  $\{1, x, x^2\}$ , and  $y = c_1 + xc_2 + x^2c_3$ .  $G_4$  is the homogeneity symmetry to indicate that the equation is autonomous and  $G_5, G_6, G_7$  form the  $sl(2, R)$  subalgebra.

**Proposition 1** *The autonomous integrating factors for (11) are given by  $y$  and  $y''$ . Using  $y''$  as the integrating factor and integrating the resulting expression,  $y''y''' = 0$ , once leads to  $\frac{1}{2}y''^2 = k$ , where  $k$  is a constant of integration. For  $k = 0$  and  $k \neq 0$  we have the  $sl(2, R)$  subalgebra and the algebra is  $sl(3, R) : 2A_1 \oplus_s \{sl(2, R) \oplus A_1\} \oplus 2A_1$ . It is also a well-known result that all equations of maximal symmetry in the form  $y^{(n)} = 0$  have the  $sl(2, R)$  element (see [13] and references therein). Otherwise if  $y'' = k$  is treated as a function we have  $G_1 = \partial_y, G_2 = x\partial_y, G_3 = \partial_x, G_4 = x\partial_x + 2y\partial_y$ . The algebra in this case is  $A_{4,9}^1 : A_2 \oplus_s 2A_1$ .*

**Proposition 2** *If  $y$  is an integrating factor of the autonomous equation of maximal symmetry  $y^{(n)} = 0$  then the integral obtained using this integrating factor will always have the  $sl(2, R)$  subalgebra.*

- In the case of equation (11) using  $y$  as an integrating factor and integrating the equation obtained once as before gives  $yy'' - \frac{1}{2}y'^2 = k$ . This equation has the  $sl(2, R)$  subalgebra. Infact the integrated equation can be written as  $(y^{1/2})'' = \frac{k}{(y^{1/2})^3}$  which is in the form of the Ermakov-Pinney equation [7, 8]. A point transformation  $v = y^{1/2}$  leads to the equation  $v'' = k/v^3$ . In this case  $k = 0$  retains the second order differential equation and  $k \neq 0$  gives the point symmetries  $G_1 = \partial_x, G_2 = 2x\partial_x + v\partial_v, G_3 = x^2\partial_x + xv\partial_v$ .
- The Ermakov-Pinney equation possesses the three-element algebra of Lie point symmetries  $sl(2, R)$  which is characteristic of all scalar ordinary differential equations of maximal symmetry (see [10, 12]).

**Proposition 3** *The equation of maximal symmetry in the form  $y^{(n)} = 0$  will have as one of its autonomous integrating factors  $y$  iff  $n$  is odd.*

As an example we can easily verify that  $y''' = 0, y^{(v)} = 0$  and  $y^{(vi)} = 0$  will all have  $y$  as one of the integrating factors. In the case where  $n$  is even  $y$  does not appear as one of the integrating factors.

## 4 Occurrence of $sl(2, R)$ integrals

For the purposes of this section we shall call the fundamental integrals as those that emanate from the solution symmetries of the differential equation. The  $sl(2, R)$  integrals will be the integrals that arise as a linear combination of the fundamental integrals and possess the  $sl(2, R)$  subalgebra. For example, the integrating factors for (11) are  $1, x, \frac{1}{2}x^2$ , from which we observe that there are also the coefficients of the solution symmetries of the differential equation. Associated with each of these are the integrals shown below:

$$\begin{array}{ll} \frac{1}{2}x^2y''' = 0 & I_1 = \frac{1}{2}x^2y'' - xy' + y \\ x.y''' = 0 & I_2 = xy'' - y' \\ 1.y''' = 0 & \longrightarrow I_3 = y'' \end{array} \quad (12)$$

In Flessas *et al* [14, 15] the numbering of the integrals is according to the one given above.

We consider the fifth order-ordinary differential equation,

$$y^v = 0, \quad (13)$$

with autonomous integrating factors  $y, y''$  and  $y^{iv}$  for illustrative purposes. Multiplying (13) by  $y$  and integrating the resulting equation gives

$$yy^{iv} - y'y''' + \frac{1}{2}y''^2 = J. \quad (14)$$

- The integral in (14) for  $J \neq 0$  has the point symmetries

$$G_1 = \partial_x \quad (15)$$

$$G_2 = x\partial_x + 2y\partial_y \quad (16)$$

$$G_3 = x^2\partial_x + 4xy\partial_y. \quad (17)$$

In the case that  $J = 0$  we just obtain the four point symmetries which are  $G_1 = \partial_x, G_2 = x\partial_x, G_3 = y\partial_y$  and  $G_4 = x^2\partial_x + 4xy\partial_y$ .

It is noted that the integral obtained using the integrating factor  $y$  always has  $sl(2, R)$  subalgebra according to proposition 2.

Equation (13) has a basis of solutions  $1, x, x^2, x^3, x^4$  so that each of these give us the solution symmetries,

$$G_1 = \partial_x$$

$$G_2 = x\partial_x$$

$$G_3 = x^2\partial_x$$

$$G_3 = x^3\partial_x$$

$$G_3 = x^4\partial_x.$$

Associated with these solution symmetries are the fundamental integrals

$$\begin{array}{ll} x^4 y^v = 0 & I_0 = \frac{1}{24} x^4 y^{iv} - \frac{1}{6} x^3 y''' + \frac{1}{2} x^2 y'' - xy' + y \\ x^3 y^v = 0 & I_1 = \frac{1}{6} x^3 y^{iv} - \frac{1}{2} x^2 y''' + xy'' - y' \\ x^2 y^v = 0 & I_2 = \frac{1}{2} x^2 y^{iv} - xy''' + y'' \\ xy^v = 0 & I_3 = xy^{iv} - y''' \\ 1y^v = 0 & \longrightarrow I_4 = y^{iv}. \end{array} \quad (18)$$

**Proposition 4** For the fifth order equation,  $y^v = 0$ , the autonomous integral emanating from the integrating factor  $y$  appearing in equation (14) can be obtained from the linear combination  $J = I_0 I_4 - I_1 I_3 + \frac{1}{2} I_2^2$  where the  $I_i, i = 0, 1, 2, 3, 4$  are as given in (18) above.

We now consider the case for which  $n = 9$ , that is, the equation which was mentioned but not treated in [13]. In this case the equation takes the form

$$y^{ix} = 0. \quad (19)$$

The autonomous integrating factors corresponding to (19) are  $y, y'', y^{iv}, y^{vi}$  and  $y^{viii}$ . If we use  $y$  in (19) as the integrating factor and integrating the subsequent equation we obtain

$$yy^{viii} - y'y^{vii} + y''y^{vi} - y'''y^v + \frac{1}{2}y^{iv^2} = J. \quad (20)$$

The integral in (20) has Lie-point symmetries



$$\begin{aligned}
G_1 &= \partial_x \\
G_2 &= x\partial_x + 4y\partial_y \\
G_3 &= x^2\partial_x + 8xy\partial_y
\end{aligned}$$

which form the  $sl(2, R)$  subalgebra as expected. The rest of the integrals corresponding to the other integrating factors are given as follows:

$$\begin{array}{ccc}
y''y^{ix} & y''y^{viii} - y'''y^{vii} + y''''y^{vi} - \frac{1}{2}y^{v^2} \\
y'''y^{ix} & y'''y^{viii} - y^vy^{vii} + \frac{1}{2}y^{vi^2} \\
y^{vi}y^{ix} & y^{vi}y^{viii} - \frac{1}{2}y^{vii^2} \\
y^{viii}y^{ix} & \longrightarrow \frac{1}{2}y^{viii^2}.
\end{array} \quad (21)$$

Moreover it is important to note that the higher the order the more tedious the expressions become for the fundamental integrals hence the need for a more compact generating function for the fundamental integrals. In the case of the 9th order ordinary differential equation we list these just to illustrate the point as follows:

$$\begin{array}{ll}
x^8y^{ix} & I_0 = x^8y^{viii} - 8x^7y^{vii} + 56x^6y^{vi} - 336x^5y^v + \\
& 1680x^4y^{iv} - 6720x^3y''' + 20160x^2y'' - 40320xy' + 40320y \\
x^7y^{ix} & I_1 = x^7y^{viii} - 7x^6y^{vii} + 42x^5y^{vi} - 210x^4y^v + \\
& 840x^3y^{iv} - 2520x^2y''' + 5040xy'' - 5040y' \\
x^6y^{ix} & I_2 = x^6y^{viii} - 6x^5y^{vii} + 30x^4y^{vi} - 120x^3y^v + \\
& 360x^2y^{iv} - 720xy''' + 720y'' \\
x^5y^{ix} & I_3 = x^5y^{viii} - 5x^4y^{vii} + 20x^3y^{vi} - 60x^2y^v + \\
& 120xy^{iv} - 120y''' \\
x^4y^{ix} & I_4 = x^4y^{viii} - 4x^3y^{vii} + 12x^2y^{vi} - 24xy^v + 24y^{iv} \\
x^3y^{ix} & I_5 = x^3y^{viii} - 3x^2y^{vii} + 6xy^{vi} - 6y^v \\
x^2y^{ix} & I_6 = x^2y^{viii} - 2xy^{vii} + 2y^{vi} \\
1y^{ix} & \longrightarrow I_7 = y^{viii}.
\end{array} \quad (22)$$

This then makes the construction of  $J$  using the fundamental integrals a daunting exercise as the order of the equation increases. However one can verify that the integral  $J = yy^{viii} - y'y^{vii} + y''y^{vi} - y'''y^v + \frac{1}{2}y^{iv^2}$  can also be obtained as a linear combination of the  $I_i$ 's with  $i = 0, 1, 2, 3, 4, 5, 6, 7$  above.

## 5 Conclusion

We have discussed some aspects on the integrability of ordinary differential equations. The discussion extended to the relationship between the basis set of solutions of an  $n$ th order ordinary differential equation of order  $n$ , its solution symmetries and associated fundamental integrals. Of interest was the integral constructed from these fundamental integrals. As the order of the equation increases, particularly, from order  $n \geq 9$  the construction of the  $sl(2, R)$  integral from the fundamental integrals would involve very long expressions and hence a need for a construction of a more compact generating function to generate all such integrals. Otherwise the same  $sl(2, R)$  integral can be obtained by direct integration from the original equation after multiplying it by the autonomous integrating factor  $y$ . We also observe that the  $sl(2, R)$  integral only occurs in  $y^{(n)} = 0$  if  $n$  is odd and  $n \geq 3$ . This discussion needs to be extended to all other equations of maximal symmetry and not necessarily with the simplest case considered here and to extend it to generalised symmetries.

## 6 Acknowledgments

The author thanks the NRF and the Durban University of Technology for their support.

## References

- [1] S Lie *Differentialgleichungen*, Reprinted by Chelsea, New York, 1967.
- [2] L Hsu & N Kamran, Classification of second order ordinary differential equations admitting Lie groups of fibre-preserving point symmetries *Proc. London Math. Society*, **58**, 387-416 (1989).
- [3] G. W Bluman & S Kumei (1989) *Symmetries and Differential Equations*, Springer-Verlag, New York, 1989.
- [4] E. S Cheb-Terrab and A.D.Roche, Integrating factors for second order ordinary differential equations, *J. Sym. Computation*, **27**, 501-519 (1999).
- [5] B Abraham-Shrauner, Hidden Symmetries, First Integrals and reduction of order of nonlinear ordinary differential equations *J. Non. Math. Physics*, **9**, Sup 2, 1-9 (2002).
- [6] P.G.L Leach & S.E Bouquet, Symmetries and integrating factors *J. Non. Math. Physics* **9**, Sup 2, 73-91 (2002).
- [7] V. Ermakov, Second order differential equations. Conditions of complete integrability,(translated by A O Harin), *Univ. Izvestia Kiev Ser III*, **9**, 1-25 (1880).

- [8] E. Pinney, The nonlinear differential equation  $y''(x) + p(x)y + cy^{-3} = 0$  *Proc. Amer. Math. Society*, **1**, 681 (1950).
- [9] A. K Head, LIE, a PC program for Lie analysis of differential equations, *Comp. Phys. Communication*, **77**, 241-248 (1993).
- [10] F. M Mahomed & P. G. L Leach, Symmetry Lie algebras of  $n$ th order ordinary differential equations, *J. Math. Anal. Applications*, **151**, 80-107 (1990).
- [11] F. M Mahomed, Symmetry group classification of ordinary differential equations: P Survey of some results, *Math. Meth. Appl. Sciences*, **30**, 1995-2012 (2007).
- [12] S Moyo & P. G. L Leach, Exceptional properties of second and third order ordinary differential equations of maximal symmetry, *J. Math. Anal. Applications*, **252**, 840-865 (2000).
- [13] S Moyo & P.G.L Leach, Symmetry properties of autonomous integrating factors, *Symme., Integrab. Geom. Meth. Applications (SIGMA)*, **1**, paper **023** (2005).
- [14] G.P Flessas, K.S Govinder & P.G.L Leach, Remarks on the symmetry Lie algebras of first integrals of scalar third order ordinary differential equations with maximal symmetry, *Bulletin of the Greek Mathematical Society*, **36**, 63-79 (1994).
- [15] G.P Flessas, K.S Govinder & P.G. L Leach PGL, Characterisation of the algebraic properties of first integrals of scalar ordinary differential equations of maximal symmetry, *J. Math. Anal. Applications*, **212**, 349-374 (1997).
- [16] P.G.L Leach, K.S Govinder and B Abraham-Shrauner, (1999), Symmetries of first integrals and their associated differential equations, *J. Math. Anal. Applications*, **Vol 235**, 58-83 (1999).

# MAC solution for a rectangular membrane

Igor Neygebauer

November 25, 2008

## Abstract

A rectangular membrane with fixed boundary conditions under applied a transversal force has the solution with singularity. That is the Green's function of this problem. But if a string is considered, which could be taken as a strip from the above membrane under the same force then the Green's function has not a singularity. The MAC solution for the membrane will be considered. This solution transforms the above Green's function with singularity into the MAC function which does not have a singularity. The obtained MAC solution for membrane corresponds to the Green's function of a string.

KEY WORDS: membrane, Laplace equation, MAC solution.

## 1 Introduction

Membranes are considered in biology, chemistry and physics. Information about experiments and theories is presented in Internet by the companies Lockheed, Volvo, U.S. Army Research Office etc. A number of journals and problems concerning membrane theories are presented in references (1) - (12). We can conclude that the membrane problem is important and it is under consideration of many research groups.

The Green's function method is an important approach in membrane theory, it is widely used in micromechanics and nanomechanics too (4), (9).

The mathematical aspect of the mechanical membrane is considered in this paper. The MAC solution of the stated problem will be considered instead of the strong or weak solutions (1), which are usually under consideration. It will be shown that the strong and the weak solutions are not physically acceptable in the case when one part of the boundary conditions is given at the point inside the domain which a membrane is occupied. The MAC model of the membrane will be obtained, which solution is called MAC solution. If the classical equation of the membrane under small deformation is a wave equation then the MAC equation is an integro-differential equation. The conformal mapping is used to create the MAC Green's function and the method of superposition is applied to create the MAC model.

## 2 Circular elastic membrane under point load

### 2.1 Statement of the membrane problem

The potential energy of the initially plane elastic membrane is

$$U = T_0 \cdot \left( \int \int_D (\sqrt{1 + u_x^2 + u_y^2} - 1) dx dy \right); \quad (1)$$

where membrane lies in plane  $(x, y)$  in its natural state,  $T_0$  is its tension on a unit of length,  $u(x, y)$  - transversal displacement of the point  $(x, y)$  of the initially plane membrane,  $D$  - domain of the plane  $(x, y)$  occupied by a membrane, external forces are not applied. Apply the following boundary conditions:

$$u|_{\partial D} = 0, \quad (2)$$

$$u(a, b) = u_0 \neq 0, \quad (3)$$

$\partial D$  is the boundary of the boundary of  $D$  without a point  $(a, b)$ , which is an internal point in  $D$ . We will consider the linearized equations of the membrane with the elastic potential energy:

$$U = \frac{T_0}{2} \cdot \left( \int \int_D (u_x^2 + u_y^2) dx dy \right); \quad (4)$$

where it is supposed that  $|u_x| \ll 1$  and  $|u_y| \ll 1$ . The membrane equation in this case of small strains is the 2D Laplace equation:

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \quad (5)$$

### 2.2 Three types of solutions of the membrane problem

The Laplace equation may have the following three different types of solutions:

- Strong solution - it satisfies the equation and continuously the boundary conditions;
- Generalized solution of the 1st kind - it is a limit of the sequence of strong solutions;
- Generalized solution of the 2nd kind (MAC solution) - it is a function which is connected with Laplace equation and some additional conditions. In this case we will create a MAC model of the problem under consideration.

Consider these three types of solutions for a circular membrane of radius  $R$ .

### 2.3 Strong solution of nonlinear membrane problem

The potential energy for a nonlinear symmetrical problem is

$$U = T_0 \cdot \left( \int \int_D (\sqrt{1 + u_r^2} - 1) \cdot r \cdot dr \right); \quad (6)$$

Then the differential equation for a circular membrane is

$$\frac{d}{dr} \left( \frac{u_r \cdot r}{\sqrt{1 + u_r^2}} \right) = 0. \quad (7)$$

The boundary conditions are

$$u(0) = u_0, u(R) = 0. \quad (8)$$

The general solution of this ordinary differential equation (7) has the form

$$u = A \cdot \text{Arch}\left(\frac{r}{A}\right) + B, \quad (9)$$

where  $A, B$  are arbitrary constants. It follows from the boundary condition

$$u(0) = u_0, \quad (10)$$

that a constant  $A = 0$ . Then a constant  $B$  must satisfy:

$$B = u_0 \neq 0, \quad (11)$$

and

$$B = 0. \quad (12)$$

So we came to the conclusion that the strong classical nonlinear solution of the problem does not exist.

### 2.4 Strong solution of linear membrane problem

Consider now the linearized membrane problem. In this case we have the elastic energy in the form

$$U = \frac{T_0}{2} \cdot \left( \int \int_D u_r^2 \cdot r \cdot dr \right). \quad (13)$$

In the axisymmetrical case the differential equation is the Laplace equation, which is

$$\frac{1}{r} \cdot \frac{\partial}{\partial r} \left( r \cdot \frac{\partial u}{\partial r} \right) = 0 \quad (14)$$

The boundary conditions are as before in nonlinear case Eqs.(8):

$$u(0) = u_0 \neq 0, u(R) = 0. \quad (15)$$

Then the general solution of the Eq.(14) is

$$u = A \cdot \ln(r) + B; \quad (16)$$

$A, B$  - arbitrary constants.

To satisfy boundary conditions we obtain

$$u(0) = u_0 = B, A = 0; u(R) = 0 = B. \quad (17)$$

It means that we have a contradiction as a constant  $B$  can not be zero and nonzero simultaneously. So we can conclude that the classical symmetrical strong solution of linear membrane problem does not exist.

## 2.5 Weak solution of linear membrane problem

We will call weak solution the generalized solution of the 1st kind which is obtained as a limit of the sequence of strong solutions as follows.

Consider linear membrane equation for a ring:

$$0 < \epsilon \leq r \leq R; \quad (18)$$

Here  $\epsilon$  is a small radius of a hole near the origin. The Laplace equation in this symmetric case is

$$\frac{1}{r} \cdot \frac{\partial}{\partial r} \left( r \cdot \frac{\partial u}{\partial r} \right) = 0 \quad (19)$$

We apply the boundary conditions

$$u(\epsilon) = u_0 \neq 0, u(R) = 0. \quad (20)$$

If  $\epsilon$  will tend to zero then the stated problem will remind us about the problem for a circle Eqs(14),(15). The solution of the given boundary value problem Eqs.(18),(19),(20) is

$$u = \frac{u_0}{\ln(\epsilon)} \cdot \ln(r) \quad (21)$$

where the external radius  $R$  is taken as  $R = 1$  that does not create any restrictions. If the internal radius

$$\epsilon \rightarrow 0 \quad (22)$$

then the generalized solution of the 1st kind for a circle or in our notations is

$$u = u_0 \text{ for } r = \epsilon \rightarrow 0, \quad (23)$$

$$u = 0 \text{ for } 0 < r \leq 1. \quad (24)$$

This solution is not a continuous function of  $r$ . This solution we will call as a weak solution of our linearized membrane problem.

## 2.6 Weak solution of nonlinear membrane problem

It is not too difficult to see that the approach from the preceding subsection, applied to the nonlinear membrane problem, does not give any solution at all, because the limit when  $\epsilon$  tends to zero does not exist.

## 2.7 MAC solution of the linear membrane problem

To overcome the problem of nonexistence of solutions of membrane problem the MAC model will be introduced. To create the MAC model consider any diameter of a circular membrane. Then the string along this diameter is taken. Consider now the following elastic string problem. The elastic potential energy of a string is

$$U = T_0 \cdot \left( \int_0^L (\sqrt{1 + u_x^2} - 1) dx \right); \quad (25)$$

We consider a stretched string which lies along the  $x$  axis in its natural state from  $x = 0$  to  $x = L$ ,  $T_0$  is a tension of a string;  $u(x)$  is transversal displacement of a point  $x$  in plain  $(x, u)$ ;  $L$  is the length of a string.

The boundary conditions of the string are

$$u(0) = 0; u(L) = 0; \quad (26)$$

$$u(a) = u_0 \neq 0, \quad (27)$$

where  $a$  is an internal point of a string,  $0 < a < L$ .

The string equation is:

$$\frac{\partial^2 u}{\partial x^2} = 0. \quad (28)$$

The solutions of the linear and nonlinear string problems coincide. That solution is

$$u(x) = u_0 \cdot \frac{x}{a}, \quad 0 \leq x \leq a; \quad (29)$$

$$u(x) = u_0 \cdot \frac{L - x}{L - a}, \quad a \leq x \leq L. \quad (30)$$

We take  $a = \frac{L}{2}$  in the obtained solution(29),(30) and because of the symmetry of the membrane problem the MAC solution of the membrane problem (14),(15) is defined as

$$u(r) = u_0 \cdot (1 - r). \quad (31)$$

It is easily to see that the introduced MAC solution Eq.(31) corresponds to the experiments with membranes. Anybody can check it at home without scientific laboratories. The strong solution of this problem does not exist and the weak solution Eqs.(23),(24) is a nonsense.

The MAC solution will be used to obtain the MAC Green's function for rectangular membrane in the next section.



### 3 MAC solution for rectangular membrane

#### 3.1 Statement of the problem

To obtain the MAC Green's function for rectangle consider the linear membrane problem for a rectangle. The rectangle occupies the domain  $D : -K_1 \leq x \leq K_1, 0 \leq y \leq K_2$ .

The linearized equation of the membrane is the Laplace equation Eq.(5) or

$$\Delta u = 0. \quad (32)$$

The boundary conditions are

$$u|_{\partial D} = 0, \quad (33)$$

$$u(a, b) = u_0 \neq 0, \quad (34)$$

where the point  $(a, b)$  is an internal point of the rectangle and  $\partial D$  is the boundary of the rectangle.

#### 3.2 Conformal mapping

The rectangle  $-K \leq x \leq K_1, 0 \leq y \leq K_2$  in the complex plane  $z = x + i \cdot y$  is conformal mapped into the circle  $x_1^2 + y_1^2 \leq 1$  in the complex plane  $z_1 = x_1 + i \cdot y_1$ :

$$z_1 = \frac{sn(z; k) - sn(d; k)}{sn(z; k) - \overline{sn(d; k)}}, \quad (35)$$

where  $0 < k < 1$  is given constant and  $d = a + i \cdot b = (a, b)$  is an inner point in rectangle;

$sn(z, k)$  is elliptical sinus;

$$K_1 = sn(1; k), \quad (36)$$

$$K_2 = sn\left(\frac{1}{k}; k\right) - K_1. \quad (37)$$

The point  $(a, b)$  in the  $z$ -plane is mapped into the origin in the  $z_1$ -plane. The unknown displacement  $u$  will then depend on the independent variables  $x_1, y_1$ . We have the following boundary value problem:

Domain  $D_1$  is a circle  $x_1^2 + y_1^2 \leq 1$ .

The equation is again the Laplace equation Eq.(5):

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial y_1^2} = 0. \quad (38)$$

The boundary conditions are

$$u|_{x_1^2 + y_1^2 = 1} = 0, \quad (39)$$

$$u(0, 0) = u_0. \quad (40)$$

The problem Eqs.(38),(39),(40) is the same problem which was considered in the preceding section. As solution of this problem we take the MAC solution Eq.(31). And finally we obtain the MAC solution of the membrane problem in the form:

$$u(x, y) = u_0 \cdot (1 - |z_1|), \quad (41)$$

where  $z_1$  is obtained from conformal mapping of rectangle into the unit circle:

$$z_1 = \frac{sn(z; k) - sn(d; k)}{sn(z; k) - \overline{sn(d; k)}}, \quad (42)$$

### 3.3 MAC Green's function

Transversal force acting at the point  $d = (a, b)$  of rectangle is:

$$P = \oint_{\partial D} \frac{\partial u}{\partial n} ds, \quad (43)$$

The function  $P(u_0, a, b)$  depends on the point  $(a, b)$  and the applied displacement  $u_0$ . If the applied point load  $P$  is given then we can solve the equation

$$P(u_0, a, b) = P \quad (44)$$

and find

$$u_0 = u_0(P, a, b) = P \cdot f(a, b) = P \cdot f(d). \quad (45)$$

If we put  $u_0$  from the Eq.(45) into the Eq.(41) then we obtain the MAC solution in the form of the MAC Green's function for a rectangular membrane. So we have

$$u(x, y) = P \cdot f(d) \cdot (1 - |z_1|) = P \cdot M(z, d), \quad (46)$$

where a point  $z = (x, y)$  and the MAC Green's function of rectangular membrane is

$$M(z, d) = f(d) \cdot (1 - |\frac{sn(z; k) - sn(d; k)}{sn(z; k) - \overline{sn(d; k)}}|). \quad (47)$$

## 4 Integro-differential equation for MAC membrane model

Principle of superposition allows to write the new membrane equation in the following form of integro-differential equation:

$$u(x, y, t) = \iint_{\Omega} M(z, d) \cdot \rho(d) \cdot \frac{\partial^2 u}{\partial t^2}(d, t) dD, \quad (48)$$

where  $M(z, d)$  - is the MAC Green's function of the domain  $D$ ,  
 $u(x, y, t)$  is the transversal displacement of the point  $d$  of the membrane,  
 $\rho(d)$  - mass-density of the membrane per unit area at a point  $d$ ,  
 $t$  - is time.

## 5 Conclusion

- The strong solution of nonlinear and linear circular membrane problem does not exist,
- The generalized solution of the 1st kind (weak solution) does not exist in nonlinear membrane problem,
- The generalized solution of the 1st kind (weak solution) exists in linear membrane problem:  
it satisfies the Laplace equation and does not satisfy the boundary conditions continuously,
- The generalized solution of the second kind (MAC solution) and MAC Green function are introduced as solutions of the MAC membrane model: they satisfy the boundary conditions and does not satisfy the Laplace equation.

The obtained integro-differential equation is suggested to use in membrane problems of more complicated type.

## 6 References

1. H.R.Clark,Elastic membrane equation in bounded and unbounded domains,*EJQTDE*,11,1-21(2002).
2. Q.Guo-zhen,Solution for free vibration problem of membrane with unequal tension in two directions,*Applied Mathematics and Mechanics*,3,6,885-892(1982).
3. W.Horn,J.Sokolowski,An elastic membrane with an attached non-linear thermoelastic rod,*Int.J.Appl.Math.Comput.Sci.*,12,4,479-486(2002).
4. U.Komaragiri,M.R.Begley,J.G.Simmonds,The mechanical response of free-standing circular elastic films under point and pressure loads,*J.of Applied Mechanics*,72,2,203-212(2005).
5. A.V.Kononov,R.De Borst,A.R.M.Wolfert,Radiation emitted by a constant load moving uniformly in a circle on an elastically supported membrane,*Wave Motion*,33,4,349-357(2001).
6. L.Lidin,Circular elastic membrane loaded at concentric circle,*AIAA Journal*,13,9,1242-1245(1975).
7. X.Li,D.J.Steigmann,Point loads on a hemispherical elastic membrane,*International Journal of Non-linear Mechanics*,30,4,569-581(1995).
8. Y.Sui,Y.T.Chew,X.B.Chen,H.T.Low,Transient deformation of elastic capsules in shear flow: Effect of membrane bending stiffness,*Physical review.E,Statistical, nonlinear, and softmatter physics*,75(2),6(2007).

9. S.Li,G.Wang,*Introduction to Micromechanics and Nanomechanics*,World Scientific,New Jersey,(2008).
10. Y.Sui,Y.T.Chew,X.B.Chen,H.T.Low,Transient deformation of elastic capsules in shear flow: Effect of membrane bending stiffness,*Physical review.E,Statistical, nonlinear, and softmatter physics*,75(2),6(2007).
11. K.Yamada,K.Suzuki,M.Hongo,Fundamental study on the characteristics of a rectangular elastic membrane in supersonic flow,*Lournal of the Japan Society for Aeronautical and Space Sciences*,52,600,30-37(2004).
12. W.Zhang,Z.Yang,J.Yang,Membrane analogy of the stevens-tiersten equation for essentially thickness modes in plate quartz resonators,*IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*,55,7,1665-1668(2008).

# Multiresolution Inverse Wavelet Reconstruction from a Fourier Partial Sum

Nataniel Greene

Department of Mathematics and Computer Science  
Kingsborough Community College, CUNY  
2001 Oriental Boulevard, Brooklyn, NY 11235  
email: [ngreene.math@gmail.com](mailto:ngreene.math@gmail.com)

March 16, 2009

The Gibbs phenomenon refers to the lack of uniform convergence which occurs in many orthogonal basis approximations to piecewise smooth functions. This lack of uniform convergence manifests itself in spurious oscillations near the points of discontinuity and a low order of convergence away from the discontinuities. In previous work [11,12] we described a numerical procedure for overcoming the Gibbs phenomenon called the Inverse Wavelet Reconstruction method (IWR). The method takes the Fourier coefficients of an oscillatory partial sum and uses them to construct the wavelet coefficients of a non-oscillatory wavelet series. However, we only described the method standard wavelet series and not for the continuous wavelet transform. We also only described the method for so-called *crude wavelets*: wavelets which do not admit a scaling function, and therefore lack a multiresolution analysis. Here we describe a version of IWR based on a quadrature discretization of the continuous inverse wavelet transform. We also describe the multiresolution potential of the Inverse wavelet reconstruction method, which is the ability to decompose a function into a coarse trend plus finer details.

Key Words: Gibbs phenomenon, Fourier Series, Inverse polynomial reconstruction, Wavelets, Inverse wavelet reconstruction.

## 1 Introduction

Fourier and orthogonal polynomial series are known for their highly accurate expansions for smooth functions. In fact it is known that the more derivatives a function has, the faster the approximation will converge. However, when a function possesses jump-discontinuities the approximation will fail to converge uniformly. In addition, spurious oscillations will cause a loss of accuracy throughout the entire domain. This lack of uniform convergence is

known as the Gibbs phenomenon. Methods for post-processing approximations which suffer from the Gibbs phenomenon include the Gegenbauer reconstruction method of Gottlieb and Shu [7, 9], the method of Pade approximants due to Driscoll and Fornberg [2], the method of spectral mollifiers due to Gottlieb and Tadmor [8] and Tadmor and Tanner [21, 22], the Inverse polynomial reconstruction (IPR) method of Shizgal and Jung [14,15,16,19,20], and the Freund polynomial reconstruction method of Gelb and Tanner [6]. These reconstruction methods can be combined with an effective method for edge-detection developed by Gelb and Tadmor [3, 4, 5], to yield an exponentially accurate reconstruction of the original function. In previous work we described a numerical method for overcoming the Gibbs phenomenon following the work of Shizgal and Jung, called the Inverse Wavelet Reconstruction method (IWR) [10,11,12]. The IWR method was shown numerically to perform at least as well as inverse polynomial reconstruction.

However, our main justification for choosing wavelets over other orthogonal or non-orthogonal bases, lies in ability of wavelets to decompose a function into a coarse trend and finer details. This decomposition is essentially what is known as a multiresolution analysis. Our previous papers on the IWR method focused on so-called *crude wavelets*: wavelets which do not admit a scaling function and which therefore do not offer a multiresolution analysis. The present paper addresses the notion of multiresolution reconstruction from a Fourier partial sum. Our previous papers described the method using standard wavelet series approximations. The present paper also addresses how the method can be implemented using the continuous wavelet transform (CWT). The CWT is truncated and discretized using a uniform Trapezoidal rule or a Gauss quadrature grid and a non-standard wavelet series approximation is obtained.

We begin with a brief review of the essential definitions of wavelets which we will need. Recall that a wavelet is a function  $\psi \in L^2(\mathbf{R})$  satisfying:

$$\int_{-\infty}^{\infty} \psi(x) dx = 0 \quad (1)$$

and

$$\int_{-\infty}^{\infty} \frac{|\Psi(\xi)|}{|\xi|} d\xi < \infty, \quad (2)$$

where  $\Psi$  here is the Fourier transform of  $\psi$ . To avoid a conflict of notations we will use a capital letter for the Fourier transform and a hat above the letter for a Fourier coefficient. The function  $\psi$  is known as a mother wavelet or an analyzing wavelet since any function  $f \in L^2(\mathbf{R})$  can be expressed as a continuous sum of translations and dilations involving  $\psi$  according to the continuous wavelet transform. The continuous wavelet transform (CWT) is given by

$$F_{\psi}(b, a) = |a|^{-1/2} \int_{-\infty}^{\infty} f(x) \overline{\psi\left(\frac{x-b}{a}\right)} dx$$

and the inverse CWT is given by

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{F_\psi(b, a)}{a^2} \psi_{b,a}(x) da db \quad (3)$$

where

$$\psi_{b,a}(x) = |a|^{-1/2} \overline{\psi\left(\frac{x-b}{a}\right)}$$

and

$$C_\psi = \int_{-\infty}^{\infty} \frac{|\Psi(\xi)|}{|\xi|} d\xi.$$

A discrete wavelet series is given for fixed constants  $a_0, b_0$  by

$$f(x) = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \psi_{m,l}(x) \quad (4)$$

where the discrete wavelet series coefficients are given by

$$d_{m,l} = F_\psi\left(\frac{b_0 l}{a_0^m}, \frac{1}{a_0^m}\right)$$

and  $\psi_{m,l}(x)$  in this case has a slightly different meaning:

$$\psi_{m,l}(x) = a_0^{m/2} \psi(a_0^m x - b_0 l).$$

In general the wavelet functions  $\psi_{m,l}$  do not constitute an orthonormal basis. Instead they constitute what is known as a frame. The family of functions  $\psi_{m,l}$  is a frame if

$$A \|f\|^2 \leq \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |\langle f, \psi_{m,l} \rangle| \leq B \|f\|^2$$

for positive constants  $A$  and  $B$ . When  $A = B$  the functions  $\psi_{j,k}$  are known as a tight frame. For more information see Daubechies [1]. The inverse wavelet reconstruction method makes use of wavelet bases and frames, as well as the CWT.

Much of the present paper focuses on wavelets which are known to have what is known as a multiresolution analysis. A scaling function, sometimes called a father wavelet, is function  $\phi(x)$  which satisfies the two-scale relations:

$$\phi(x) = \sum_{l=-\infty}^{\infty} h_l \phi(2x - l)$$

and

$$\psi(x) = \sum_{l=-\infty}^{\infty} g_l \phi(2x - l).$$

The functions  $\phi_{m,l}(x) = 2^{m/2}\phi(2^m x - l)$  are orthonormal to each other and they are also orthonormal to  $\psi_{m,l}(x) = 2^{m/2}\psi(2^m x - l)$ . The functions  $\phi$  and  $\psi$  allow for a decomposition of a function into the sum of a coarse approximation and finer details as follows:

$$f(x) = \sum_{l=-\infty}^{\infty} c_{M_0,l} \phi_{M_0,l}(x) + \sum_{m=M_0}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \psi_{m,l}(x)$$

which is truncated as

$$f_{M,L}(x) \simeq \sum_{l=-L}^L c_{M_0,l} \phi_{M_0,l}(x) + \sum_{m=M_0}^{M-1} \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x).$$

The  $c_{M_0,l}$  are known as approximation coefficients for the function at resolution scale  $M_0$  and the  $d_{m,l}$  are detail coefficients for the finer scales. This particular decomposition of a function into trend and details is known as a multiresolution analysis.

## 2 The Multiresolution Inverse Wavelet Reconstruction Method

### 2.1 Inverse Wavelet Reconstruction

We begin with a Fourier partial sum and assume that the function  $f(x)$  can also be expressed as a discrete wavelet series:

$$f(x) = \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \psi_{m,l}(x)$$

where

$$d_{m,l} = \int_{-\infty}^{\infty} f(x) \overline{\psi_{m,l}(x)} dx$$

and

$$\psi_{m,l}(x) = a_0^{-m/2} \psi(a_0^m x - b_0 l).$$

We now derive a formula expressing the Fourier coefficients in terms of wavelet coefficients.

$$\begin{aligned} \hat{f}(n) &= \frac{1}{2} \int_{-1}^1 f(x) e^{-i\pi n x} dx \\ &= \frac{1}{2} \int_{-1}^1 \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \psi_{m,l}(x) e^{-i\pi n x} dx \\ &= \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \left( \frac{1}{2} \int_{-1}^1 \psi_{m,l}(x) e^{-i\pi n x} dx \right) \\ &= \sum_{m=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \hat{\psi}_{m,l}(n) \end{aligned} \tag{5}$$



The inverse wavelet reconstruction method is obtained by truncating the doubly infinite sum described above and solving for the wavelet coefficients. We suggest solving the following system of equations.

$$\hat{f}(n) = \sum_{m=-M}^M \sum_{l=-L}^L d_{m,l} \hat{\psi}_{m,l}(n) \quad (6)$$

for the wavelet coefficients  $d_{m,l}$  where  $n = -N \dots N$  and  $N = 2ML + M + L$ . One then computes the IWR approximation

$$S_{M,L}f(x) = \sum_{m=-M}^M \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x). \quad (7)$$

The terms  $\hat{\psi}_{m,l}(n)$  are the  $n$ th Fourier coefficients of  $\psi_{m,l}(x)$ . Since we are solving a system of  $2N + 1$  equations in  $(2M + 1)(2L + 1)$  unknown wavelet coefficients, in order for the system to be invertible we must have  $2N + 1 = (2M + 1)(2L + 1)$ , or  $N = 2ML + M + L$ . Numerical experiments indicate that the invertibility condition can be relaxed and fewer wavelet coefficients can be computed if one solves the overdetermined system in a least squares sense using Matlab's backslash operator. This reconstruction is valid for both crude and non-crude wavelets.

## 2.2 Multiresolution IWR: Wavelet and Scaling Function Approach

Here we describe the IWR method for wavelets which allow for a multiresolution reconstruction. A formal derivation of the method proceeds again as follows. Write

$$\begin{aligned} \hat{f}(n) &= \frac{1}{2} \int_{-1}^1 f(x) e^{-i\pi n x} dx \\ &= \frac{1}{2} \int_{-1}^1 \left( \sum_{l=-\infty}^{\infty} c_{M_0,l} \phi_{M_0,l}(x) + \sum_{m=M_0}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \psi_{m,l}(x) \right) e^{-i\pi n x} dx \\ &= \sum_{l=-\infty}^{\infty} c_{M_0,l} \left( \frac{1}{2} \int_{-1}^1 \phi_{M_0,l}(x) e^{-i\pi n x} dx \right) \\ &\quad + \sum_{m=M_0}^{\infty} \sum_{l=-\infty}^{\infty} d_{m,l} \left( \frac{1}{2} \int_{-1}^1 \psi_{m,l}(x) e^{-i\pi n x} dx \right) \\ &= \sum_{l=-\infty}^{\infty} c_{M_0,l} \hat{\phi}_{M_0,l}(n) + \sum_{m=M_0}^{\infty} \sum_{l=-\infty}^{\infty} c_{m,l} \hat{\psi}_{m,l}(n) \end{aligned} \quad (8)$$

The IWR method in this case is obtained by truncating the sums and solving the system of equations for the approximation and detail coefficients

$$\hat{f}(n) = \sum_{l=-L}^L c_{M_0,l} \hat{\phi}_{M_0,l}(n) + \sum_{m=M_0}^{M-1} \sum_{l=-L}^L d_{m,l} \hat{\psi}_{m,l}(n) \quad (9)$$

where  $n = -N \dots N$ . Numerical results again indicate that the system can be solved in a least squares sense and one need not be concerned with meeting an invertibility condition exactly. One then constructs the multiresolution IWR approximation

$$S_{M,L}^{M_0} f(x) = A_L^{M_0}(x) + D_{M,L}^{M_0}(x)$$

where the coarse trend is given by

$$A_L^{M_0}(x) = \sum_{l=-L}^L c_{M_0,l} \phi_{M_0,l}(x)$$

and the finer details are given by

$$D_{M,L}^{M_0}(x) = \sum_{m=M_0}^{M-1} \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x).$$

### 2.3 Multiresolution IWR: Pure Wavelet Approach

Next we describe an alternative multiresolution reconstruction for non-crude wavelets which does not make explicit use of the scaling function  $\phi$ . It is simply a modification of the original IWR approximation. We write

$$S_{M,L}^{M_0} f(x) = A_{M,L}^{M_0}(x) + D_{M,L}^{M_0}(x)$$

where the coarse trend is given by

$$A_{M,L}^{M_0}(x) = \sum_{m=-M}^{M_0-1} \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x)$$

and the finer details are given by

$$D_{M,L}^{M_0}(x) = \sum_{m=M_0}^M \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x).$$

The derivation is simply to begin with the original IWR reconstruction and write

$$\begin{aligned} S_{M,L} f(x) &= \sum_{m=-M}^M \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x) \\ &= \sum_{m=-M}^{M_0-1} \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x) + \sum_{m=M_0}^M \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x) \\ &= A_{M,L}^{M_0}(x) + D_{M,L}^{M_0}(x) \end{aligned}$$

The Pure Wavelet Approach will only provide a good decomposition into trend and details for wavelets which admit a scaling function, even though the method does not make explicit use of the scaling function. This is because the two approaches are approximately equivalent, that is:

$$A_L^{M_0}(x) \simeq A_{M,L}^{M_0}(x).$$

In fact,

$$A_L^{M_0}(x) = \sum_{l=-L}^L c_{M_0,l} \phi_{M_0,l}(x)$$

but also

$$\begin{aligned} A_L^{M_0}(x) &= \sum_{m=-\infty}^{M_0-1} \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x) \\ &= \sum_{m=-\infty}^{-M-1} \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x) + \sum_{m=-M}^{M_0-1} \sum_{l=-L}^L d_{m,l} \psi_{m,l}(x) \\ &= A_L^{-M}(x) + A_{M,L}^{M_0}(x). \end{aligned}$$

Since  $A_L^{-M}(x)$  is the tail of a convergent series,  $A_L^{-M}(x) \rightarrow 0$  as  $M \rightarrow \infty$  since

### 3 A Version of IWR Based on the Continuous Wavelet Transform

Finally, we describe here a related method for implementing IWR which is based on a quadrature discretization of the inverse continuous wavelet transform. It produces results which are comparable to the standard wavelet series approach described previously. Our chief reason for considering it is allows us to consider any wavelet family as a candidate for use in reconstruction.

The derivation is as follows.

$$\begin{aligned} \hat{f}(n) &= \frac{1}{2} \int_{-1}^1 f(x) e^{-i\pi n x} dx \\ &= \frac{1}{2} \int_{-1}^1 \left( \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{F_\psi(b,a)}{a^2} \psi\left(\frac{x-b}{a}\right) dadb \right) e^{-i\pi n x} dx \\ &= \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{F_\psi(b,a)}{a^2} \left( \frac{1}{2} \int_{-1}^1 \psi\left(\frac{x-b}{a}\right) e^{-i\pi n x} dx \right) dadb \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{C_\psi} \frac{F_\psi(b,a)}{a^2} \hat{\psi}_{b,a}(n) dadb. \end{aligned} \tag{10}$$

Note that in this section we use  $\hat{\psi}_{b,a}(n)$  to denote the  $n$ th Fourier coefficient of  $\psi\left(\frac{x-b}{a}\right)$ .

Truncate the limits of integration and write

$$\hat{f}(n) \simeq \int_{-B}^B \int_{-A}^A \frac{1}{C_\psi} \frac{F_\psi(b, a)}{a^2} \hat{\psi}_{b,a}(n) da db. \quad (11)$$

Next, discretize the integrations using a quadrature rule such as the trapezoidal rule or Gauss quadrature. The resulting equations become

$$\hat{f}(n) \simeq \sum_{l=0}^L \sum_{m=0}^M \frac{1}{C_\psi} \frac{F_\psi(b_l, a_m)}{a_m^2} \hat{\psi}_{b_l, a_l}(n) \quad (12)$$

which can be expressed in simpler form by letting  $d_{m,l} = \frac{1}{C_\psi} \frac{F_\psi(b_l, a_m)}{a_m^2}$  and writing

$$\hat{f}(n) \simeq \sum_{m=0}^M \sum_{l=0}^L d_{m,l} \hat{\psi}_{b_l, a_l}(n). \quad (13)$$

Finally, solve the system of equations, possibly in a least squares sense, for the coefficients  $d_{m,l}$  and construct the nonstandard wavelet series approximation:

$$f(x) \simeq \sum_{m=0}^M \sum_{l=0}^L d_{m,l} \psi\left(\frac{x - b_l}{a_m}\right). \quad (14)$$

## 4 Numerical Results

Numerical experiments show that the IWR method based on a quadrature discretization of the CWT yields uniformly converging approximations for all wavelet families we have tested. We display the results of the Haar, Shannon, and Mexican hat wavelets in the figures below. The Haar and Shannon systems are some of the earliest examples of what later became wavelets. The Haar wavelet is apparently able to resolve an interior discontinuity without prior knowledge as to its location, as well as the endpoints. Therefore the Haar wavelet is able to provide a global reconstruction, albeit a low order accuracy one. The low order of convergence is due to the fact that the wavelet is not well-localized in Fourier space. Shannon wavelets will resolve the Gibbs phenomenon at the endpoints of the interval and Mexican hat wavelets will do it even better.

Numerical experiments also show that the Multiresolution IWR method is able to decompose a function into a coarse trend plus finer details, given a function's Fourier coefficients. We provide numerical examples using the both the Pure Wavelet Approach to reconstruction and the Wavelet and Scaling function Approach described above. We implemented the inverse wavelet reconstruction method by computing Fourier coefficients for the test functions in question using trapezoidal rule quadrature. This makes our Fourier series a pseudospectral series. For the method to be accurate we compute the coefficients  $\hat{\psi}_{m,l}(n)$  using the same trapezoidal rule quadrature.

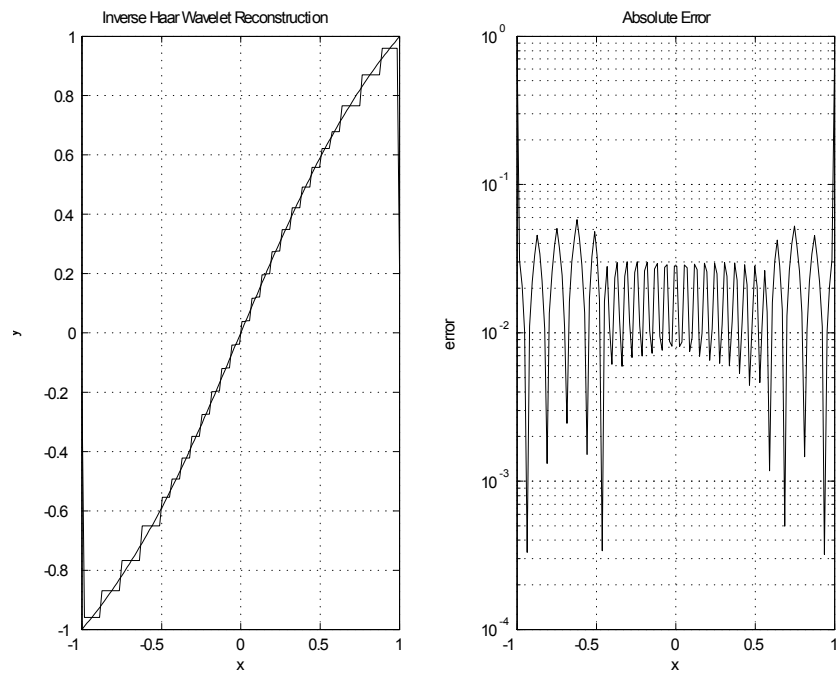


Figure 1: Inverse Haar Reconstruction of  $\frac{4}{\pi} \tan^{-1} x$  based on 64 Fourier coefficients using the wavelet series approach.

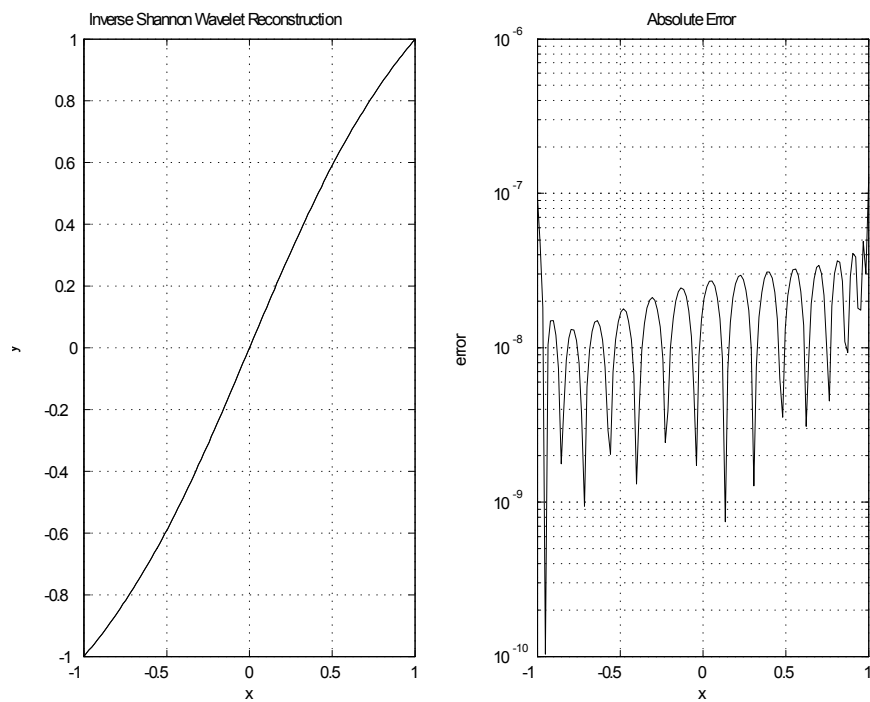


Figure 2: Inverse Shannon reconstruction of  $\frac{4}{\pi} \tan^{-1} x$  using the CWT approach from 128 Fourier coefficients. The limits of integration are  $A = B = 20$  and 64 Gauss-Legendre quadrature nodes in  $a$  and  $b$  are employed.

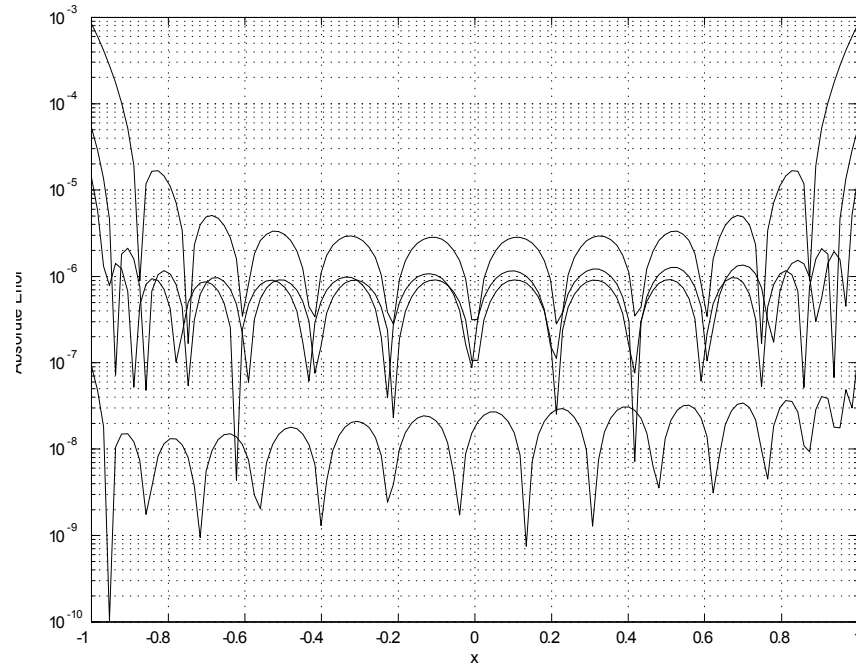


Figure 3: Converge of Inverse Shannon reconstruction of  $\frac{4}{\pi} \tan^{-1} x$  using the CWT approach from 16, 32, 64, and 128 Fourier coefficients. The limits of integration are  $A = B = 20$  and the number of Gauss-Legendre quadrature nodes in  $a$  and  $b$  are equal to the number of Fourier coefficients.

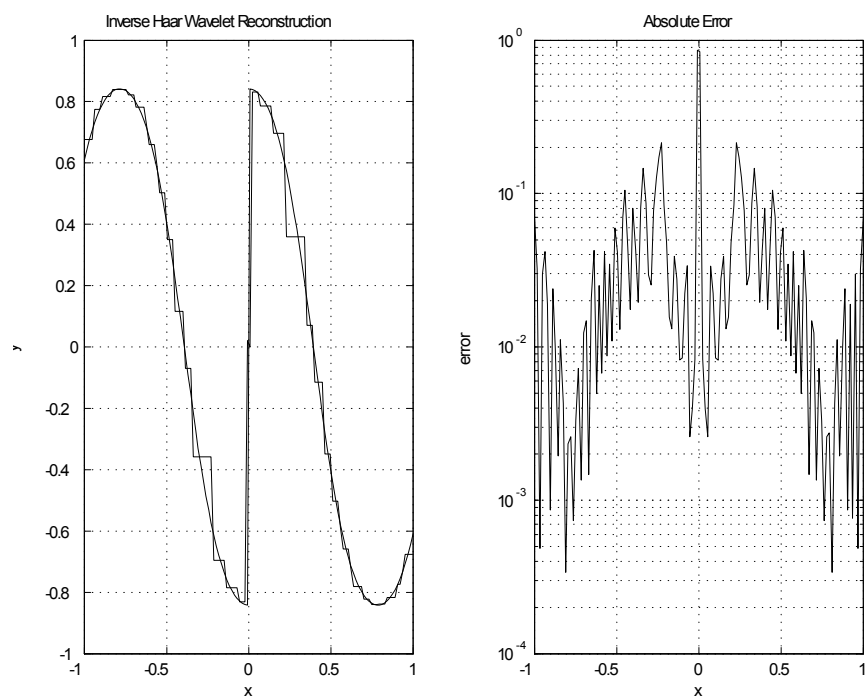


Figure 4: Inverse Haar reconstruction of  $\sin(\cos(4x))\operatorname{sgn}(x)$  utilizing the CWT approach with  $A = B = 20$ , and Gauss-Legendre quadrature with 32 quadrature nodes in  $a$  and 32 quadrature nodes in  $b$ . The method resolves both endpoint and interior discontinuities, albeit with low accuracy.



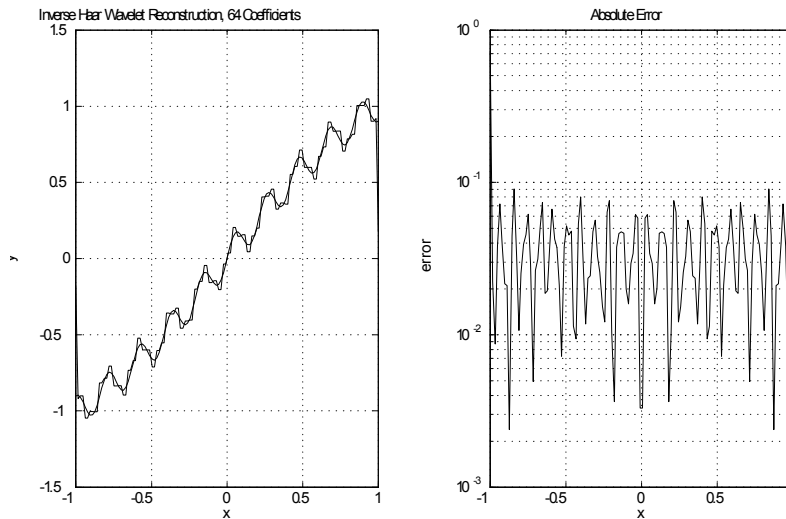


Figure 5: Inverse Haar wavelet reconstruction using the pure wavelet approach of the function  $f(x) = \frac{4}{\pi} \tan^{-1} x + \frac{1}{10} \sin 30x$  based on 64 Fourier coefficients.

We consider the test function  $f(x) = \frac{4}{\pi} \tan^{-1}(x) + \frac{1}{10} \sin(30x)$  to observe the ability of various wavelets to extract the trend, mostly due to  $\tan^{-1}(x)$ , and smaller scale oscillations due to  $\frac{1}{10} \sin(30x)$ . We display here the results of the Haar wavelet: the simplest wavelet which admits a multiresolution analysis. We also consider the piecewise linear Franklin wavelets generated by the second order splines.

## 5 Conclusions

Numerical results indicate that the IWR method based on the continuous wavelet transform works as well as the IWR method based on a standard wavelet series. It yields an accurate and rapidly converging approximation. The multiresolution IWR methods allows one to separate out from the Fourier coefficients a course trend and finer details regardless of any Gibbs phenomenon. Work in progress includes a numerical study of the multiresolution IWR method for a wider variety of wavelets, multiresolution reconstruction from series other than Fourier series, a two-dimensional implementation and a proof of convergence.

## References

- [1] I. Daubechies, Ten Lectures on Wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, 1992.

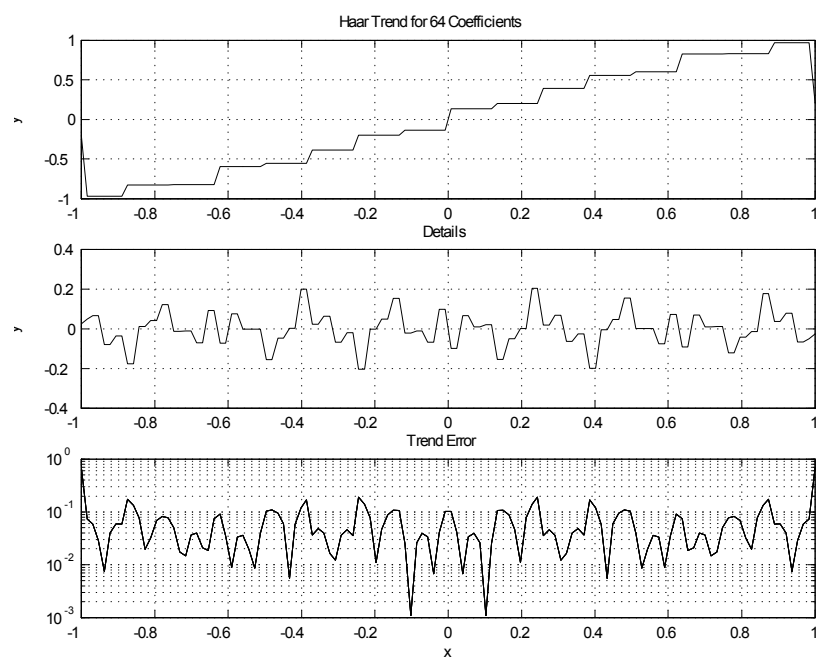
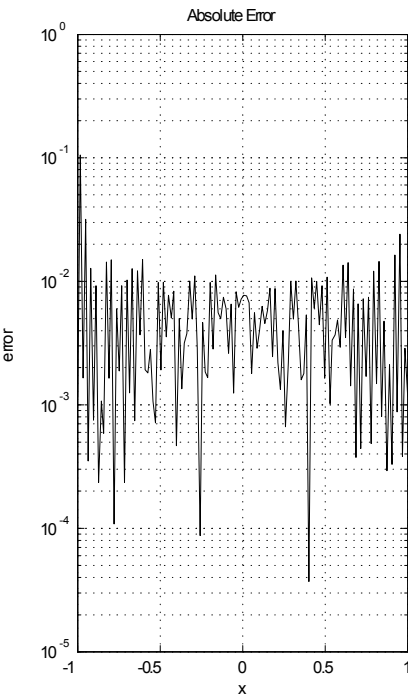
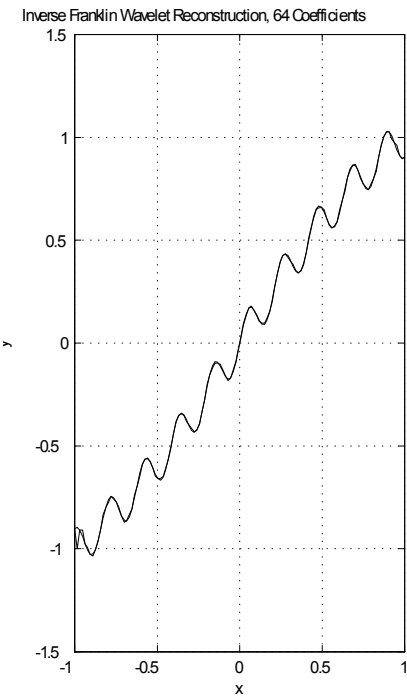


Figure 6: Inverse Haar decomposition of  $f(x) = \frac{4}{\pi} \tan^{-1} x + \frac{1}{10} \sin 30x$  into a trend and details using the pure wavelet approach.



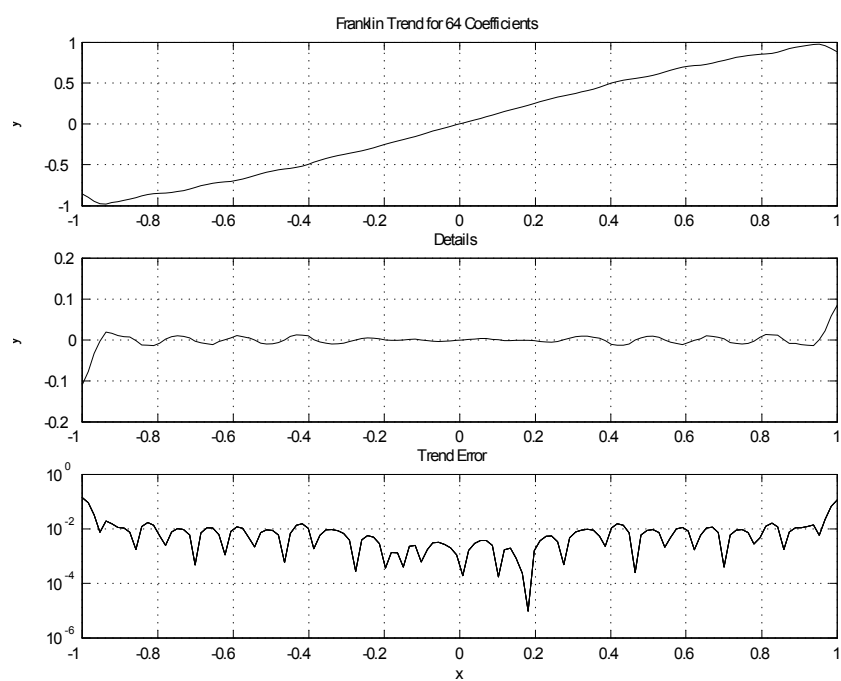


Figure 7: Inverse Franklin decomposition of  $f(x) = \frac{4}{\pi} \tan^{-1} x + \frac{1}{10} \sin 30x$  into a trend and details.

- [2] T. A. Driscoll, and B. Fornberg, A Pade-based Algorithm for Overcoming the Gibbs Phenomenon, *Numerical Algorithms*, 26, 2001, pp. 77-92.
- [3] A. Gelb, and E. Tadmor, Detection of Edges in Spectral Data, *Appl. Comp. Harmonic Anal.*, 7, 1999, pp. 101-135.
- [4] A. Gelb, and E. Tadmor, Detection of Edges in Spectral Data II. Nonlinear Enhancement, *SIAM J. Numer. Anal.*, Vol. 38, No. 4, 2000, pp. 1389-1408.
- [5] A. Gelb, and E. Tadmor, Spectral Reconstruction of Piecewise Smooth Functions from their Discrete Data, *Mathematical Modeling and Numerical Analysis*, 36:2, 2002, pp. 155-175.
- [6] A. Gelb, and J. Tanner, Robust Reprojection Methods for the Resolution of the Gibbs Phenomenon., *Applied Computational and Harmonic Analysis*, Vol. 20, 1, 2006, pp. 3-25.
- [7] D. Gottlieb, C.-W. Shu, A. Solomonoff, and H. Vandeven, On the Gibbs's Phenomenon I: Recovering Exponential Accuracy from the Fourier Partial Sum of a Nonperiodic Analytic Function, *J. Comput. Appl. Math.*, 43, 1992, pp. 81-92 .
- [8] D. Gottlieb, and E. Tadmor, Recovering Pointwise Values of Discontinuous Data within Spectral Accuracy, in: *Progress and Supercomputing in Computational Fluid Dynamics*, Proceedings of a 1984 U.S.-Israel Workshop, Progress in Scientific Computing, Vol. 6 (E. M. Murman and S. S. Abarbanel eds.), Birkhauser, Boston, 1985, pp. 357-375.
- [9] D. Gottlieb and C.-W. Shu, On the Gibbs Phenomenon and its Resolution, *SIAM Rev.*, Vol. 39, 1997, pp. 644-668.
- [10] N. Greene, *On the Recovery of Piecewise Smooth Functions from their Integral Transforms and Spectral Data*, Ph.D. Dissertation, SUNY Stony Brook, 2004.
- [11] N. Greene, A Wavelet-based Method for Overcoming the Gibbs Phenomenon, in: *Recent Advances on Applied Mathematics: Proceedings of the American Conference on Applied Mathematics*, Cambridge, Massachusetts, March 24-26, 2008, pp. 408-412.
- [12] N. Greene, Inverse Wavelet Reconstruction for Resolving the Gibbs Phenomenon, *International Journal of Circuits, Systems and Signal Processing*, Issue 2, Vol. 2, 2008, pp. 73-77.
- [13] J.-H. Jung, and B. D. Shizgal, Generalization of the Inverse Polynomial Reconstruction Method in the Resolution of the Gibbs Phenomenon, *J. Comp. Appl. Math.*, v. 172, n.1, 2004, pp.131-151.

- [14] J.-H. Jung, and B. D. Shizgal, Inverse polynomial reconstruction of two dimensional Fourier images, *J. Scientific Computing*, v.25, n.3, 2005, pp. 367-399.
- [15] J.-H. Jung, and B. D. Shizgal, On the numerical convergence with the inverse polynomial reconstruction method for the resolution of the Gibbs phenomenon, *J. Computational Physics*, 224, 2007, pp. 477-488.
- [16] B. D. Shizgal and J.-H. Jung, Towards the resolution of the Gibbs Phenomena, *J. Comput. Appl. Math.*, 16, 2003, pp. 41-65.
- [17] R. Pasquetti, On Inverse methods for the Resolution of the Gibbs Phenomenon, *Journal of Computational and Applied Mathematics*, Vol. 170, no. 2, 2004, pp. 305-315.
- [18] E. Tadmor and J. Tanner, Adaptive Mollifiers - High Resolution Recovery of Piecewise Smooth Data from its Spectral Information, *J. Foundations of Comp. Math.* 2, 2002, 155-189.
- [19] E. Tadmor and J. Tanner, Adaptive Filters for Piecewise Smooth Spectral Data, *IMA J. Numerical Anal.*, Vol. 25, 4, 2005, pp. 635-647.

# THE CALCULATION OF AXISYMMETRIC DUCT GEOMETRIES FOR INCOMPRESSIBLE, ROTATIONAL FLOW USING AN INTEGRAL FORMULA BASED ON GREEN'S THEOREM.

Vasos Pavlika

Harrow School of Computer Science  
University of Westminster, School of Computer Science, Harrow, Middlesex, UK  
[V.L.Pavlika@westminster.ac.uk](mailto:V.L.Pavlika@westminster.ac.uk)

Key Words: Rotational Incompressible flow, Green's function and Integral Formula.

## ABSTRACT.

In this paper a numerical algorithm is described for solving the boundary value problem associated with axisymmetric, inviscid, incompressible, rotational to obtain duct wall geometries from prescribed wall velocity distributions. The governing equations are formulated in terms of the stream function  $\psi(x,y)$  and the function  $\phi(x,y)$  as independent variables where for irrotational flow  $\phi(x,y)$  can be recognized as the velocity potential function, for rotational flow  $\phi(x,y)$  ceases being the velocity potential function but does remain orthogonal to the stream lines. The numerical method uses an integral formula on a uniform mesh. The technique described is capable of tackling the so-called inverse problem where the velocity wall distributions are prescribed from which the duct wall shape is calculated, as well as the direct problem where the velocity distribution on the duct walls are calculated from prescribed duct wall shapes. Numerical results for the case of the Dirichlet boundary conditions are given.

## 1. INTRODUCTION

Designers of ducts require numerical techniques for calculating wall shapes from a prescribed velocity distribution. The objective of the prescribed velocity is typically to avoid boundary layer separation. At inlet a velocity is prescribed to allow for a vorticity to be present calculated from  $\underline{\omega} = \nabla \wedge \underline{v}$  where the  $\wedge$  denotes the usual cross product of vectors,  $\underline{\omega}$  is the vorticity vector and  $\underline{v}$  the velocity vector respectively.

The present paper describes a numerical algorithm for solving the boundary value problem that arises when the independent variables are  $\phi$  and  $\psi$ , where  $\phi$  may be identified as the velocity potential function (for irrotational flow only), for flow with vorticity,  $\phi$  ceases being the velocity potential function but does remain orthogonal to  $\psi$  which may be identified as the stream function. The dependent variable  $y$ , is the radial coordinate and  $x$  is the axial coordinate. The numerical technique is based on an integral formula using finite difference on a uniform mesh.

## 2. THE DESIGN PLANE.

As shown in Pavlika [4] when the independent variables are  $\phi(x,y)$  and  $\psi(x,y)$  where the  $\phi(x,y)$  and  $\psi(x,y)$  have been previously defined it can be shown that the governing partial differential equation that the radius satisfies is given by:

$$\frac{\partial}{\partial \phi} \left( \frac{A}{B} \frac{\partial y}{\partial \phi} \right) + \frac{\partial}{\partial \psi} \left( \frac{B}{A} \frac{\partial y}{\partial \psi} \right) = 0 \quad (1)$$

with the speed calculated from

$$\frac{1}{q^2} = \frac{1}{A^2} \left( \frac{\partial y}{\partial \psi} \right)^2 + \frac{1}{B^2} \left( \frac{\partial y}{\partial \phi} \right)^2 \quad (2)$$

and completion of the physical coordinates are provided from

$$dx = \frac{B}{A} \frac{\partial y}{\partial \psi} d\phi - \frac{A}{B} \frac{\partial y}{\partial \phi} d\psi$$

where  $x$  is the axial coordinate and  $A$  and  $B$  satisfy their own first order quasi-linear hyperbolic partial differential equations with characteristics parallel to the  $\phi$  and  $\psi$  axes which maps the physical flow field into an infinite strip in the  $(\phi, \psi)$  plane. In fact the  $A$  and  $B$

$$\text{satisfy: } \frac{\partial}{\partial \phi} (\log(A)) = \frac{\eta}{q^2} B \quad (3)$$

$$\text{and } \frac{\partial}{\partial \psi} (\log(B)) = -\frac{\omega_\alpha}{q^2} A \quad (4)$$

Regarding temporarily  $\eta$ ,  $\omega_\alpha$  and  $q$  as known functions of  $\phi$  and  $\psi$  the system (3) and (4) as previously mentioned is quasi-linear hyperbolic with characteristics parallel to the  $\psi$  and  $\phi$  axes which maps the physical flow field into an infinite strip in the  $(\phi, \psi)$  plane. Bearing in mind the freedom available in the stream wise variation of  $\phi$  and the cross stream variation of  $\psi$ , suitable values of  $A$  can be prescribed along one  $\phi$  characteristic and those of  $B$  can be prescribed along one  $\psi$  characteristic.

## 3. The NUMERICAL ALGORITHM IN THE DESIGN PLANE.

Rewriting the partial differential equation that  $y$  satisfies i.e. Eq. (1) as:

$$\frac{\partial}{\partial \phi} \left( C \frac{\partial y}{\partial \phi} \right) + \frac{\partial}{\partial \psi} \left( \frac{1}{C} \frac{\partial y}{\partial \psi} \right) = c \quad (5)$$

where  $C = \frac{A}{B}$ , for problems posed in the design plane

$c=0$ , the value of  $C$  will vary depending on whether the flow field is irrotational or swirl free etc. Eq. (5) will be re-written as a Poisson equation that is as:

$$\nabla^2 y = \frac{c}{C} + \left(1 - \frac{1}{C^2}\right) \frac{\partial^2 y}{\partial \psi^2} - \left( \frac{\partial}{\partial \phi} \log_e |C| \right) \frac{\partial y}{\partial \phi} - \frac{1}{C} \frac{\partial}{\partial \psi} \left( \frac{1}{C} \right) \frac{\partial y}{\partial \psi} \quad (6)$$

where  $\nabla^2$  is the usual Laplacian operator in rectangular

coordinates so that  $\nabla^2 y = g\left(\frac{\partial^2 y}{\partial \psi^2}, \frac{\partial y}{\partial \phi}, \frac{\partial y}{\partial \psi}, C, c\right)$

where  $g$  is a function of the arguments shown as defined by expression (6).

#### 4. SOLUTION BASED ON AN INTEGRAL FORMULA USING GREEN'S THEOREM.

Here the method of solution is derived using an integral formula. Commencing with the generalized form of Green's theorem for the self adjoint elliptic operator  $E(t)$  in normal form given by:

$$\iint_R v E(t) - t E^{(A)}(v) d\phi d\psi = \oint_C t \frac{\partial v}{\partial n} - v \frac{\partial t}{\partial n} ds$$

where  $t = y^2$ ,  $E(u) = E^{(A)}(u)$  where  $E^{(A)}(u)$  is the adjoint of  $E$  and  $v$  is the fundamental solution to the adjoint equation. In this case the adjoint equation is given by  $\nabla^2 v = 0$  and  $E(t) = g$  as defined by Eq. (6). The contour  $C$  bounding the surface  $R$  is traversed in the counter clockwise sense. For a doubly connected region introducing a singularity at the point  $(\phi_0, \psi_0)$  (inside or on the contour  $C$ ) and assuming

$v(\phi, \psi) = F(\phi, \psi) \log_e |r|$  so that the distance  $r$  is

given by:  $r = \left( (\phi - \phi_0)^2 + (\psi - \psi_0)^2 \right)^{1/2}$

with  $F(\phi, \psi)$  analytic, then it can be shown that

$$m\pi t(\phi_0, \psi_0) F(\phi_0, \psi_0) = \oint_C t \frac{\partial v}{\partial n} - v \frac{\partial t}{\partial n} ds - \iint_R v g \left( \frac{\partial^2 t}{\partial \phi^2}, \frac{\partial^2 L}{\partial \phi^2} \right) d\phi d\psi, \quad m=1, 2$$

with  $L = \log_e t$ . Now  $m = 2$  if  $(\phi_0, \psi_0)$  is within  $C$  and  $m=1$  if  $(\phi_0, \psi_0)$  is on  $C$  (the  $m=1$  case can be shown using the appropriate Plemelj formulae or by indenting the contour at  $(\phi_0, \psi_0)$ ). For the Dirichlet case of boundary condition of  $t(\phi, \psi)$  the requirement

is that  $v(\phi, \psi) = 0$  on  $C$  in addition to  $v(\phi, \psi)$  being harmonic and for the Neumann conditions on  $t$ , the requirement is that  $\frac{\partial v}{\partial n} = 0$  and  $v(\phi, \psi)$  once again

satisfying Laplace's equation. Much literature is available for the Green's function for the Laplace equation (see Williams [6]) and need not be mentioned here. Hence for the Dirichlet problem without loss of generality setting  $F(\phi, \psi) = 1 \forall \phi, \psi$  and for interior points:

$$2\pi t(\phi_0, \psi_0) = - \oint_C t \frac{\partial v_D}{\partial n} ds - \iint_R v_D g \left( \frac{\partial^2 t}{\partial \phi^2}, \frac{\partial^2 L}{\partial \phi^2} \right) d\phi d\psi \quad (7)$$

i.e., the Green's function  $v_D$  satisfies Laplace's equation on  $\partial\Omega$  where  $\partial\Omega$  is defined by:

$\phi_0 \leq \phi \leq \phi_{M+1}, \psi_0 \leq \psi \leq \psi_{N+1}$  and vanishes on  $C$ . For the Neumann problem

$$2\pi t(\phi_0, \psi_0) = - \oint_C v_N \frac{\partial t}{\partial n} ds - \iint_R v_N g \left( \frac{\partial^2 t}{\partial \phi^2}, \frac{\partial^2 L}{\partial \phi^2} \right) d\phi d\psi$$

Which give integral formulae for the square of the radius  $t$ , from which the radius  $y$  can be determined., above Green's function  $v_N$  satisfies the Laplace equation on

$\partial\Omega$  with  $\frac{\partial v_N}{\partial n}$  vanishing on  $C$ . Knowledge of the

derivatives  $\frac{\partial t}{\partial \phi}$  and  $\frac{\partial t}{\partial \psi}$  are also required for the

determination of the speed  $q$  given by Eq. (2) hence differentiating under the integral sign above with respect

to  $\phi$  and  $\psi$  gives integral formulae for both  $\frac{\partial t}{\partial \phi}$  and

$\frac{\partial t}{\partial \psi}$ , such that:

$$2\pi \frac{\partial}{\partial \psi} t(\phi_0, \psi_0) = \oint_C \frac{\partial t}{\partial \psi} \cdot \frac{\partial v_D}{\partial n} + t \frac{\partial^2 v_D}{\partial \psi \partial n} ds - \iint_R v_D \frac{\partial}{\partial \psi} \left( g \left( \frac{\partial^2 t}{\partial \phi^2}, \frac{\partial^2 L}{\partial \phi^2} \right) \right) + \frac{\partial v_D}{\partial \psi} \cdot g \left( \frac{\partial^2 t}{\partial \phi^2}, \frac{\partial^2 L}{\partial \phi^2} \right) d\phi d\psi$$

and similarly for  $\frac{\partial t}{\partial \phi}$

$$2\pi \frac{\partial}{\partial \phi} t(\phi_0, \psi_0) = \oint_C \frac{\partial t}{\partial \phi} \cdot \frac{\partial v_D}{\partial n} + t \frac{\partial^2 v_D}{\partial \phi \partial n} ds - \iint_R v_D \frac{\partial}{\partial \phi} \left( g \left( \frac{\partial^2 t}{\partial \phi^2}, \frac{\partial^2 L}{\partial \phi^2} \right) \right) + \frac{\partial v_D}{\partial \phi} \cdot g \left( \frac{\partial^2 t}{\partial \phi^2}, \frac{\partial^2 L}{\partial \phi^2} \right) d\phi d\psi$$



## 5. ITERATIVE SOLUTION.

To convert formula (7) to a system of linear algebraic equations the point  $t(\varphi_0, \psi_0)$  inside  $C$  is related to its boundary values on  $C$ . To obtain the first iterates  $t_i^{(1)}(\varphi_0, \psi_0)$ ,  $g_i^{(0)}$  is set equal to zero, so that

$$2\pi t_i^{(1)} = \sum_j^{2N+2M+4} \left( \frac{\partial v_D}{\partial n} \right)_j t_j \Delta s_j \quad i=0,1,2,\dots,2N+2M+4$$

Using the trapezoidal rule

$$\pi t_i^{(1)} = \sum_j^{2N+2M+4} \frac{1}{4} \left( \frac{\partial v_D}{\partial n} \right)_j (s_{j+1} - s_{j-1}) t_j$$

$$i=0,1,2,\dots,2N+2M+4$$

$$\Rightarrow t_i^{(1)} = \sum_j^{2N+2M+4} K(v_D, s)_j t_j, i=0,1,2,\dots,2N+2M+4$$

$$\text{Where } K(v_D, s) = \frac{1}{4\pi} \left( \frac{\partial v_D}{\partial n} \right) (s_{j+1} - s_{j-1})$$

Using this method there is a simple self-consistency check. i.e. the  $t_j$  are known upstream and downstream for  $j=0,1,2,\dots,N+1$  and  $j=N+M+3, N+M+4,\dots,2N+M+3$ , hence the first iteration may be written as:

$$\begin{bmatrix} 1-K_{N+2} & -K_{N+3} & \cdot & \cdot & -K_{2N+2M+4} \\ -K_{N+2} & 1-K_{N+3} & -K_{N+4} & \cdot & -K_{2N+2M+4} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -K_{N+2} & \cdot & \cdot & \cdot & -K_{2N+2M+4} \end{bmatrix} \begin{bmatrix} t_{N+2} \\ t_{N+3} \\ \cdot \\ \cdot \\ t_{2N+2M+4} \end{bmatrix} = \begin{bmatrix} \sum_j K_j t_j \\ \sum_j K_j t_j \\ \cdot \\ \cdot \\ \sum_j K_j t_j \end{bmatrix}$$

$$\text{so that } A^{(1)} \underline{t}^{(1)} = \underline{b}^{(1)} \quad (8)$$

where the summations on the right hand side are performed over  $j=0,1,\dots,N+1$  and  $j=M+N+3, M+N+4,\dots,2M+N+3$ . Once the first iterate  $t_j$  has been calculated the field integral containing  $g$  is then computed, where the central difference approximation to the second derivative is used, this is then introduced into the right hand side of Eq. (8) and compute the second iterate,  $\underline{t}^{(2)}$ . The procedure is repeated until some convergence criteria is satisfied e.g.  $\| \underline{t}_i^{(k)} - \underline{t}_i^{(k-1)} \|_p < \epsilon$ , where  $\epsilon$  is a constant and the  $p$  denotes the  $p$ -norm ( $p=1, 2$  or  $\infty$ ).

## 6. AXISYMMETRIC FLOW IN THE ABSENCE OF BODY FORCES.

Here numerical solutions to inviscid axisymmetric flow with constant vorticity and a swirl velocity will be derived. The axial velocity component  $u_x(y)$  at inlet will be chosen to be of the form  $u_x(y) = \alpha y + \beta$ , where  $\alpha$  and  $\beta$  are constants chosen such that  $u_x(y_1) = u_1$  and  $u_x(y_2) = u_2$  where  $y_1$  represents the inner radius and  $y_2$  the outer radius at inlet. The swirl velocity  $u_\alpha(y)$ , will be of the

form  $u_\alpha(y) = ky + \frac{l}{y}$  where the  $k$  and  $l$  are constants

with  $ky$  representing solid body rotation and  $l/y$  the so-called free vortex term respectively.

## 7. THE FLOW EQUATIONS IN THE PHYSICAL PLANE $(y, \alpha, x)$ .

Adopting cylindrical polar coordinates with  $y$  being the radial coordinate,  $\alpha$  the circumferential and  $x$  the axial coordinate, defining velocity components  $u_y$ ,  $u_\alpha$  and  $u_x$  with corresponding vorticity components  $\omega_y$ ,  $\omega_\alpha$ ,  $\omega_x$  in the direction of increasing  $y$ ,  $\alpha$  and  $x$  respectively, then the equation of motion with unit density becomes:

$$\frac{Du}{Dt} = -\nabla \cdot \underline{p} \quad (9)$$

Where  $\frac{D}{Dt}$  is the material derivative. Eq. (9) can be

written using well known vector identities as:

$$\begin{aligned} \frac{\partial u_y}{\partial t} + u_x \frac{\partial u_y}{\partial x} + u_y \frac{\partial u_y}{\partial y} - \frac{u_\alpha^2}{y} &= -\frac{\partial p}{\partial y} \\ \frac{\partial u_\alpha}{\partial t} + u_x \frac{\partial u_\alpha}{\partial x} + u_y \frac{\partial u_\alpha}{\partial y} - \frac{u_\alpha u_y}{y} &= 0 \\ \frac{\partial u_x}{\partial t} + u_x \frac{\partial u_x}{\partial x} + u_y \frac{\partial u_x}{\partial y} &= -\frac{\partial p}{\partial x} \end{aligned} \quad (10)$$

$$\text{Furthermore } \frac{\partial u}{\partial t} + (\underline{u} \cdot \nabla) \underline{u} = -\nabla \cdot \underline{p}$$

can be written (once again using an appropriate vector identity as)

$$\frac{\partial u}{\partial t} + (\underline{\omega} \wedge \underline{u}) = -\nabla \left( p + \frac{1}{2} q^2 \right). \text{ Thus}$$

for steady flow Crocco's form of the equation of motion is obtained, i.e.

$$(\underline{u} \wedge \underline{\omega}) = \nabla H \quad (11)$$

where  $H$  is the total head defined by  $H = p + \frac{1}{2}q^2$ .

Calculating the cross product on the left hand side of Eq. (11), gives

$$\begin{aligned}\frac{\partial H}{\partial y} &= u_\alpha \omega_x - u_x \omega_\alpha \\ 0 &= u_x \omega_y - u_y \omega_x \\ \frac{\partial H}{\partial x} &= u_y \omega_\alpha - u_\alpha \omega_x\end{aligned}\quad (12)$$

In addition for axisymmetric flow the vorticity,  $\underline{\omega}$  becomes

$$\begin{aligned}\underline{\omega} = \nabla \wedge \underline{u} &= \left\{ -\frac{\partial u_\alpha}{\partial x} \right\} \underline{y} + \left\{ \frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y} \right\} \underline{\alpha} \\ &+ \left\{ \frac{1}{y} \frac{\partial(yu_\alpha)}{\partial y} \right\} \underline{x}\end{aligned}\quad (13)$$

and the equation of continuity is given by

$$\nabla \cdot \underline{u} = \frac{\partial(yu_x)}{\partial x} + \frac{\partial(yu_y)}{\partial y} = 0$$

## 8. THE DESIGN PLANE COUNTERPARTS.

In order to compute numerical solutions in the design plane, expressions are required for the terms  $A$ ,  $B$  and  $\omega_\alpha$ , thus

$$\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} = -\frac{1}{y} \left( u_x \frac{\partial y}{\partial x} + u_y \right) = -q \frac{\partial}{\partial s} (\log(y))$$

$$\text{or } \eta = -\frac{q^2}{B} \frac{\partial}{\partial \phi} (\log(y)),$$

$$\text{but } \eta = \frac{q^2}{B} \frac{\partial}{\partial \phi} (\log(A))$$

thus  $Ay = f(\psi)$ , that is  $\frac{\partial \psi}{\partial n} = \frac{yq}{f(\psi)}$ . The arbitrary

function  $f(\psi)$  represents the freedom in the cross stream distribution of  $\psi$  and choosing  $f(\psi)$  to be unity everywhere  $\psi$  can be identified as the usual Stokes stream function given by

$$\frac{\partial \psi}{\partial x} = -yu_y; \frac{\partial \psi}{\partial y} = yu_x$$

Eq. (13), (circumferential component) gives

$$0 = u_x \frac{\partial(yu_\alpha)}{\partial x} + u_y \frac{\partial(yu_\alpha)}{\partial y}$$

Referring to the meridional plane figure 1, it may be deduced that

$$u_x = q \frac{\partial x}{\partial s}; u_y = q \frac{\partial y}{\partial s}$$

$$\Rightarrow \frac{\partial}{\partial s} (yu_\alpha) = 0 \Rightarrow yu_\alpha = C(\psi)$$

where  $q = \frac{ds}{dt}$ . In terms of  $C(\psi)$  the vorticity vector

(Eq. (13)) becomes

$$\begin{aligned}\underline{\omega} = \nabla \wedge \underline{u} &= \left\{ -\frac{1}{y} \frac{\partial C}{\partial x} \right\} \underline{y} + \left\{ \frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y} \right\} \underline{\alpha} \\ &+ \left\{ \frac{1}{y} \frac{\partial C}{\partial y} \right\} \underline{x} \\ &= \omega_y \underline{y} + \omega_\alpha \underline{\alpha} + \omega_x \underline{x}, \text{ by definition.}\end{aligned}$$

An expression for  $\omega_\alpha$  is required as this appears in the expression for  $B$ , so using the radial component of Eq. (12) gives

$$\omega_\alpha = \frac{u_x}{u_\alpha} \left( \frac{1}{y} \frac{\partial C}{\partial y} \right) - \frac{1}{u_x} \frac{\partial H}{\partial y}$$

using the Stokes' stream function this becomes

$$\omega_\alpha = \frac{C(\psi)}{y} \left( \frac{dC}{d\psi} \right) - y \frac{dH}{d\psi}$$

which is the required expression to be used in calculation of  $B$  according to definition (4). If far upstream the flow is assumed to be cylindrical so that all quantities are independent of  $x$ , then with unit density the equation of motion and the Stokes' Stream function give

$$u_y = 0; \frac{\partial p}{\partial x} = 0; \frac{\partial p}{\partial y} = \frac{u_\alpha^2}{y}; \frac{\partial \psi}{\partial x} = 0; \frac{\partial \psi}{\partial y} = yu_x$$

giving

$$\omega_\alpha = \frac{C(\psi)}{y} \left( \frac{dC}{d\psi} \right) - \frac{y}{2} \frac{d}{d\psi} (u_x^2 + u_\alpha^2) - \frac{u_\alpha^2}{u_x y}$$

With  $u_x(y) = \alpha y + \beta$  and  $u_\alpha(y) = ky + \frac{l}{y}$  as

previously defined. Once  $\frac{dH}{d\psi}$  has been calculated

upstream it takes this value throughout the  $(\phi, \psi)$  since as is self evident the expression is independent of  $\phi$ .

This last expression for  $\omega_\alpha$  is required in the calculation of  $B$  and numerical coupling with Eq. (1) gives the numerical solution in the design plane.

## 9. DOWNSTREAM CONDITIONS.

Downstream a cylindrical flow condition as discussed below will be prescribed. Defining the pressure function  $H(\psi)$  and the function  $C(\psi)$  as

$$H(\psi) = \frac{1}{2}(u_x^2 + u_\alpha^2) + \frac{p}{\rho} \text{ and } C(\psi) = yu_\alpha$$

for cylindrical flow radial equilibrium (from Eq. (10) radial component gives

$$\begin{aligned}\frac{1}{\rho} \frac{dp}{dy} &= \frac{u_\alpha^2}{y}, \text{ integrating gives} \\ \frac{1}{\rho} (p - p_{y\text{-inner}}) &= \int_{y\text{-inner}} \frac{u_\alpha^2}{y} dy = \int_{y\text{-inner}} \frac{C^2(\psi)}{y^3} dy \\ \Rightarrow H(\psi) &= \frac{1}{2} (u_x^2 + u_\alpha^2) + \frac{p_{y\text{-inner}}}{\rho} + \int_{y\text{-inner}} \frac{C^2(\psi)}{y^3} dy \\ \text{Now } \int_{y\text{-inner}} \frac{C^2(\psi)}{y^3} dy &= -\frac{1}{2} \int_{y\text{-inner}} C^2 d(1/y^2) \\ &= -\frac{1}{2} \left[ \frac{C^2}{y^2} - \left( \frac{C^2}{y^2} \right)_{y\text{-inner}} \right] + \frac{1}{2} \int_{y\text{-inner}} \frac{1}{y^2} \frac{dC^2}{dy} dy \\ \therefore H(\psi) &= \frac{1}{2} u_x^2 + \frac{p_{y\text{-inner}}}{\rho} + \frac{1}{2} (u_\alpha^2)_{y\text{-inner}} \\ &\quad + \int_{\psi=0} \frac{1}{y^2} \frac{dC^2}{d\psi} d\psi\end{aligned}$$

Suppose  $u_{x,1} = u_{x,1}(\psi)$  and  $u_{\alpha,1} = u_{\alpha,1}(\psi)$ , where the subscript 1 denotes upstream conditions, then  $u_{x,2} = u_{x,2}(\psi)$  and  $u_{\alpha,2} = u_{\alpha,2}(\psi)$  are required as functions of  $\psi$ , where the subscript 2 similarly denoting downstream conditions, so that

$$\begin{aligned}\frac{1}{2} u_{x,2}^2 &= H(\psi) - \frac{p_{2,inner}}{\rho} - \frac{1}{2} (u_{\alpha,2}^2)_{inner} \\ &\quad - \frac{1}{2} \int_{\psi=0} \frac{1}{y_1^2} \frac{dC^2}{d\psi} d\psi\end{aligned} \quad (14)$$

and

$$\int_{\psi=0} \frac{d\psi}{u_{x,2}} d\psi = \frac{1}{2} (y_2^2 - y_{2,inner}^2)$$

thus  $C(\psi) = y_1 u_{\alpha,1} = y_2 u_{\alpha,2}$ , and (14) now gives

$$\begin{aligned}\frac{1}{2} u_{x,2}^2 &= \frac{1}{2} u_{x,1}^2 + \frac{p_{1,inner}}{\rho} - \frac{p_{2,inner}}{\rho} + \\ \frac{1}{2} ((u_{\alpha,1}^2)_{inner} - (u_{\alpha,2}^2)_{inner}) &+ \frac{1}{2} \int_{\psi=0} \left( \frac{1}{y_1^2} - \frac{1}{y_2^2} \right) d(C^2) \\ \text{or } u_{x,2}^2 &= u_{x,1}^2 + K + \int_{\psi=0} \left( \frac{1}{y_1^2} - \frac{1}{y_2^2} \right) d(C^2) \quad (15)\end{aligned}$$

where

$$\begin{aligned}K &= 2 \left( \frac{p_{1,inner}}{\rho} - \frac{p_{2,inner}}{\rho} \right) + (u_{\alpha,1}^2)_{inner} - (u_{\alpha,2}^2)_{inner} \\ \text{and } y_2^2 &= y_{2,inner}^2 + 2 \int_{\psi=0} \frac{d\psi}{u_{x,2}}\end{aligned} \quad (16)$$

with  $u_{x,2}$  in this case given by Eq. (15).

## 10. CALCULATION PROCEDURE.

The calculation of the downstream radii  $y_2(\psi)$  follow from Eq. (16) with  $u_{x,2}$  given by Eq. (15), which can be written as

$$u_{x,2}^2 = g(\psi) + K, \text{ where} \quad (17)$$

$$g(\psi) = u_{x,1}^2 + \int_{\psi=0} \left( \frac{1}{y_1^2} - \frac{1}{y_2^2} \right) \frac{d(C^2)}{d\psi} d\psi$$

In order to calculate the  $(n+1)^{th}$  iterate it is known that:

$$\begin{aligned}\frac{\partial}{\partial K} (y_{2,outer}^2) &= 2 \int_{\psi=0} \frac{\partial}{\partial K} \left( \frac{d\psi}{\sqrt{g(\psi) + K}} \right) \\ &= - \int_{\psi=0}^{\Psi} \frac{d\psi}{(u_{x,2}^3)^{(n)}}\end{aligned}$$

but

$$\left( \frac{\partial}{\partial K} (y_{2,outer}^2) \right)^{(n)} = \frac{(y_{2,outer}^2)^{(n+1)} - (y_{2,outer}^2)^{(n)}}{K^{(n+1)} - K^{(n)}} \quad (18)$$

from which as can be seen from Eq. (18) the  $K^{(n)}$  must be calculated iteratively with  $K^{(0)}=0$ . Once the  $K^{(n+1)}$  has been calculated it is introduced into Eq. (17), giving rise to a new  $(u_{x,2}^2)^{(n+1)}$  which in turn gives a new  $(y_{x,2}^2)^{(n+1)}$  from Eq. (16) and the process repeated until some convergence criteria is satisfied.

## 11. PRESCRIPTION OF WALL GEOMETRIES.

In this paper the Dirichlet boundary conditions will be prescribed on the wall boundaries so that it is the radii values,  $y$  that are given as a function of  $\varphi$  on the boundaries. The function chosen to give a  $y$  distribution is based on the hyperbolic tangent, choosing  $y(\varphi) = C \tanh(a\varphi + b) + k$  where  $C$ ,  $a$ ,  $b$  and  $k$  are constants, applying the conditions that  $y = y_u$  at  $\varphi = 0$  and  $y = y_d$  at  $\varphi = \Phi$  taking a  $\Phi + b = 3$  (arbitrary) and  $b = -3$ , so that  $\tanh(a\Phi + b) \approx 1$  and  $\tanh(b) \approx -1$ , then it follows that

$$y(\varphi) = \left( \frac{y_d - y_u}{2} \right) \tanh(a\varphi + b) + \left( \frac{y_d + y_u}{2} \right) \quad (19)$$

replacing  $\varphi$  by  $x$  in Eq. (19) gives a  $y(x)$  distribution. The inner radius is prescribed to be equal to unity in this paper (arbitrary). The geometries produced are shown in figures 2, 3 and 4 respectively

## 12. CONCLUSIONS.

As shown, geometries have been produced subject to given upstream and downstream conditions with prescribed Dirichlet boundary conditions. In this case vorticity at inlet has been specified by defining the axial velocity to be of the form  $u_x(y) = \alpha y + \beta$ , and the swirl velocity of the form  $u_\alpha(y) = ky + \frac{l}{y}$ , where the  $k$  and

$l$  are constants, defining the so-called free and forced vortex whirl respectively. The downstream conditions where such that: cylindrical flow was present, Dirichlet boundary conditions were prescribed, however the case with Neumann conditions can be accommodated using the algorithm, in addition so can the case with Robin boundary condition. Further examples of the algorithm with a combination of boundary condition is given in Pavlika (5). Additional work has been carried out by the author in which the inlet (and consequently exit) axial velocity profile is of parabolic type modeling effectively viscous flow as shown in Pavlika (5). Furthermore Hagen-Poiseuille flow has also been modeled using the techniques described in this paper. In addition to the techniques described above the author has also looked at the cases when body forces exist as well as blockage effects created by a circumferentially arranged cascade of aerofoils. It was found that at most twenty iterations were required to achieve an acceptable level of convergence when these additional features were incorporated into the model.

## 13. REFERENCES.

1. J.M. Cousins, *Special Computational problems associated with axisymmetric flow in Turbomachines*. Ph.D thesis (CNAA), 1976.
2. N. Curle, H.J. Davies, *Modern Fluid Dynamics*, van Nostrand Reinhold Company, 1971 Chapter 1.
3. M. Klier, *Aerodynamic Design of Annular Ducts*, Ph.D thesis (CNAA), 1990 Chapter 1.
4. V. Pavlika, *Vector Field Methods and the Hydrodynamic Design of Annular Ducts*, Ph.D thesis, University of North London, Chapter VI, 1995.
5. V. Pavlika, *Vector Field Methods and the Hydrodynamic Design of Annular Ducts*, Ph.D thesis, University of North London, Chapter VIII, 1995.
6. W.E. Williams, *Partial Differential Equations*, Clarendon Press, Oxford, 1980.

## 14. FIGURES.

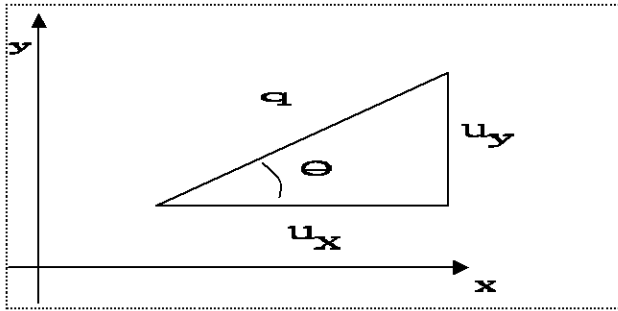


Figure 1. The meridional plane.

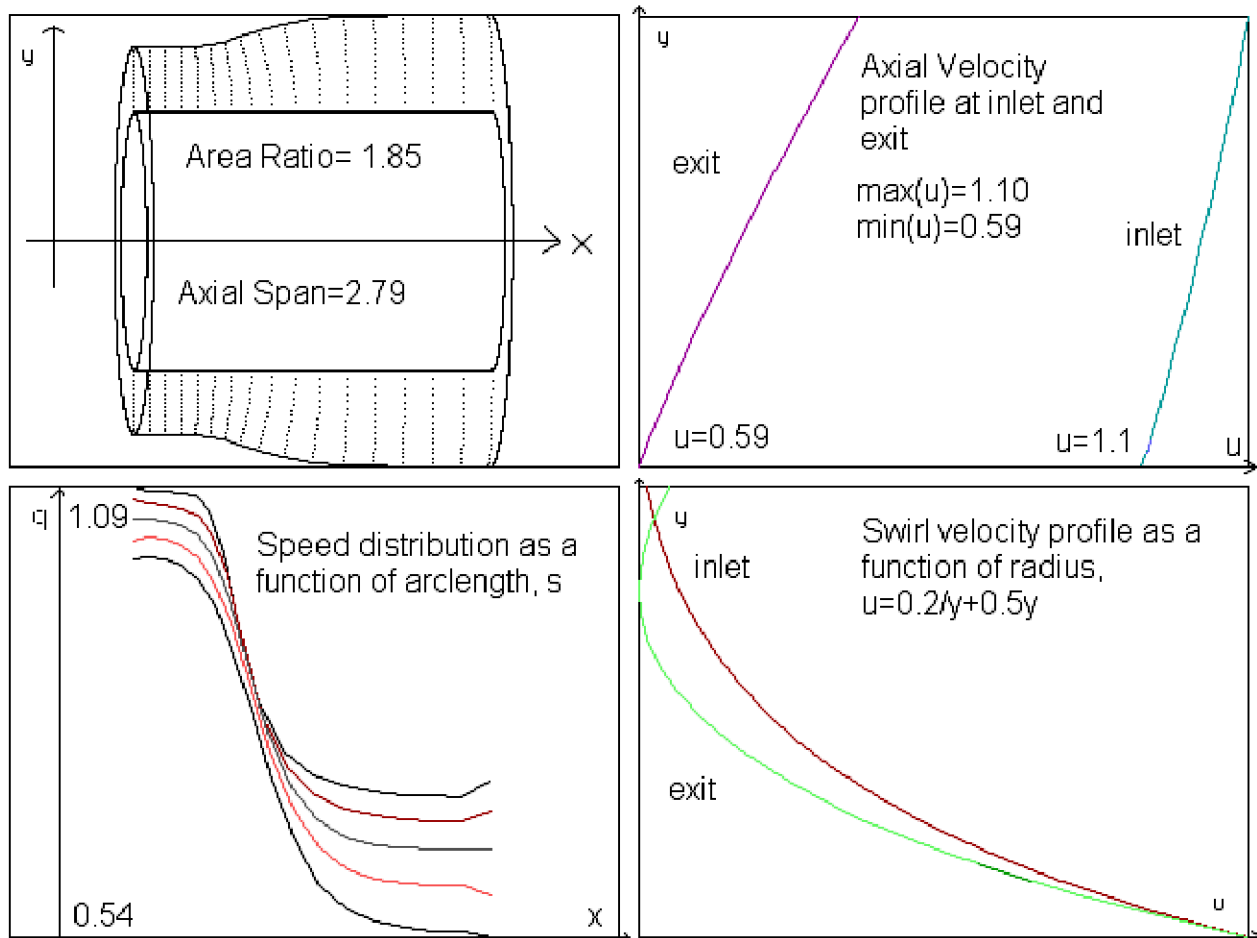
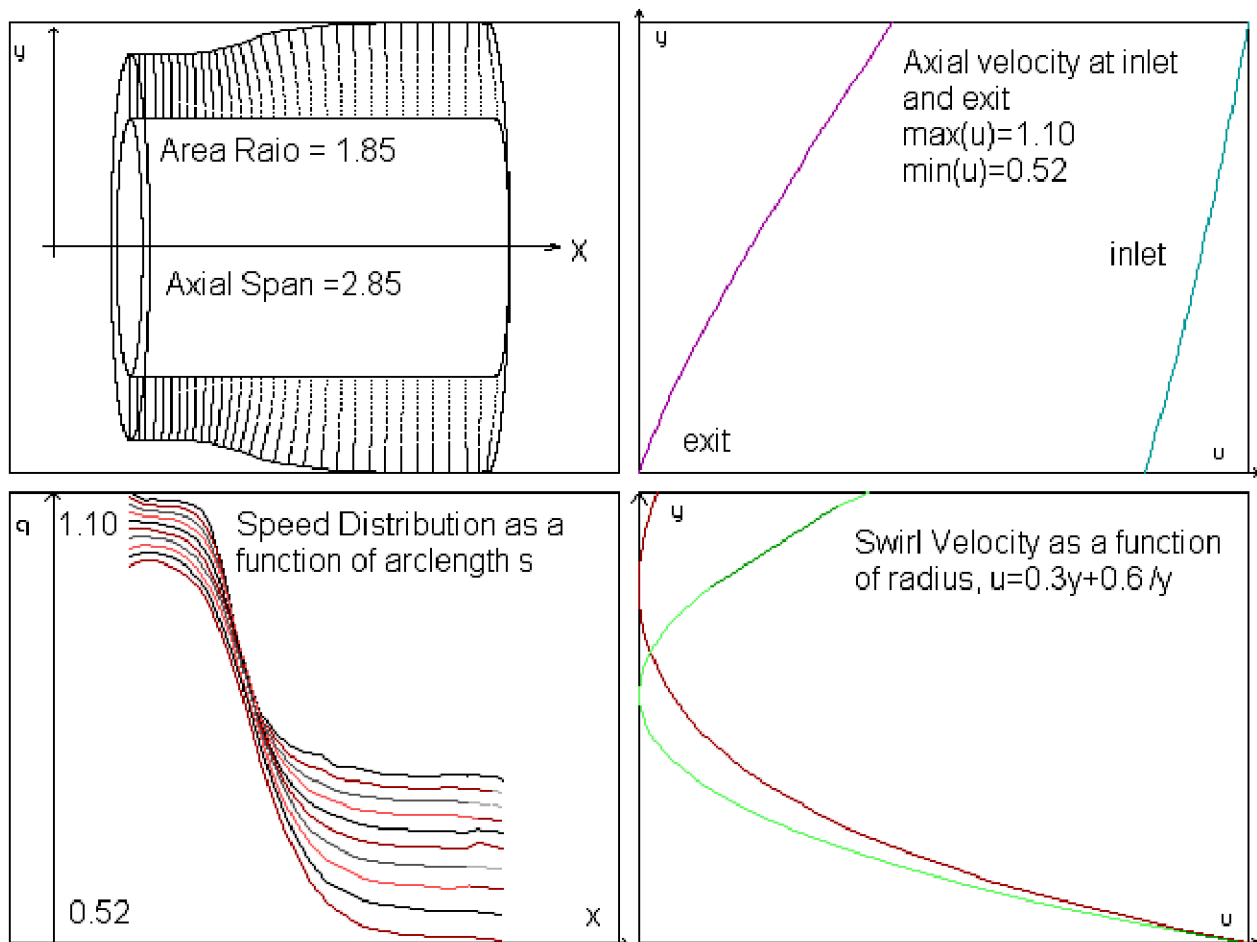
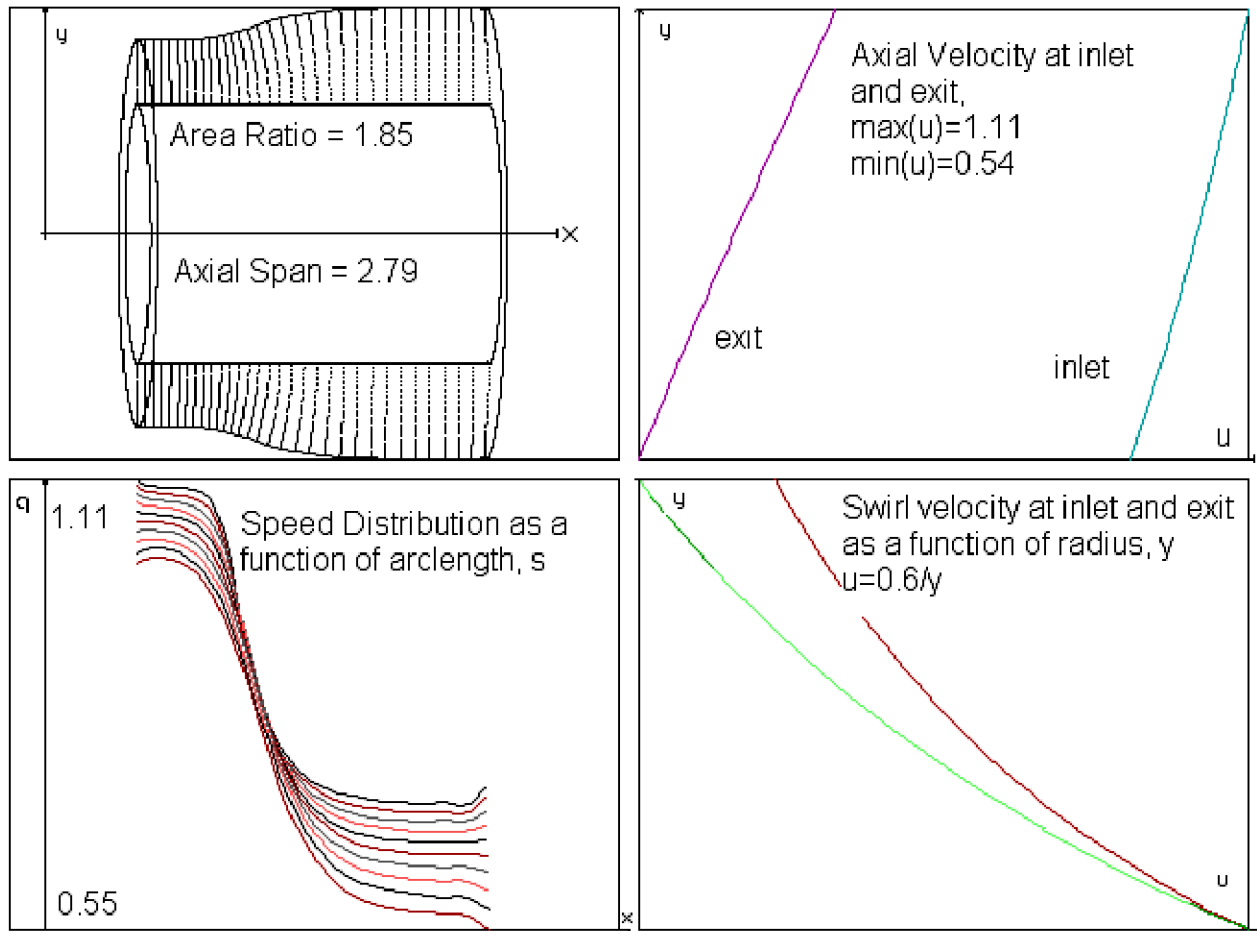


Figure 2. The geometry and speed distribution (along the top boundary) produced given a Swirl velocity  $u_\alpha = 0.5y + \frac{0.2}{y}$  and a velocity at inlet given by  $u_x(y) = \alpha y + \beta$ ,



**Figure 3.** The geometry and speed distribution (along the top boundary) produced given a Swirl velocity  $u_\alpha = 0.3y + \frac{0.6}{y}$  and an axial velocity at inlet given by  $u_x(y) = \alpha y + \beta$ ,



**Figure 4.** The geometry and speed distribution (along the top boundary) produced given a Swirl velocity  $u_\alpha = \frac{0.6}{y}$  and an axial velocity at inlet given by  $u_s(y) = \alpha y + \beta$ ,

---

**Instructions to Contributors**  
**Journal of Concrete and Applicable Mathematics**  
 A quarterly international publication of Eudoxus Press, LLC, of TN.

**Editor in Chief: George Anastassiou**  
 Department of Mathematical Sciences  
 University of Memphis  
 Memphis, TN 38152-3240, U.S.A.

**1. Manuscripts hard copies in triplicate, and in English, should be submitted to the Editor-in-Chief:**

**Prof. George A. Anastassiou**  
 Department of Mathematical Sciences  
 The University of Memphis  
 Memphis, TN 38152, USA.  
 Tel. 901.678.3144  
 e-mail: [ganastss@memphis.edu](mailto:ganastss@memphis.edu)

Authors may want to recommend an associate editor the most related to the submission to possibly handle it.

Also authors may want to submit a list of six possible referees, to be used in case we cannot find related referees by ourselves.

**2. Manuscripts should be typed using any of TEX, LaTeX, AMS-TEX, or AMS-LaTeX and according to EUDOXUS PRESS, LLC. LATEX STYLE FILE. (Click [HERE](#) to save a copy of the style file.) They should be carefully prepared in all respects. Submitted copies should be brightly printed (not dot-matrix), double spaced, in ten point type size, on one side high quality paper 8(1/2)x11 inch. Manuscripts should have generous margins on all sides and should not exceed 24 pages.**

**3. Submission is a representation that the manuscript has not been published previously in this or any other similar form and is not currently under consideration for publication elsewhere. A statement transferring from the authors (or their employers, if they hold the copyright) to Eudoxus Press, LLC, will be required before the manuscript can be accepted for publication. The Editor-in-Chief will supply the necessary forms for this transfer. Such a written transfer of copyright, which previously was assumed to be implicit in the act of submitting a manuscript, is necessary under the U.S. Copyright Law in order for the publisher to carry through the dissemination of research results and reviews as widely and effectively as possible.**



4. The paper starts with the title of the article, author's name(s) (no titles or degrees), author's affiliation(s) and e-mail addresses. The affiliation should comprise the department, institution (usually university or company), city, state (and/or nation) and mail code.

The following items, 5 and 6, should be on page no. 1 of the paper.

5. An abstract is to be provided, preferably no longer than 150 words.

6. A list of 5 key words is to be provided directly below the abstract. Key words should express the precise content of the manuscript, as they are used for indexing purposes.

The main body of the paper should begin on page no. 1, if possible.

7. All sections should be numbered with Arabic numerals (such as: 1. INTRODUCTION).

Subsections should be identified with section and subsection numbers (such as 6.1. Second-Value Subheading).

If applicable, an independent single-number system (one for each category) should be used to label all theorems, lemmas, propositions, corollaries, definitions, remarks, examples, etc. The label (such as Lemma 7) should be typed with paragraph indentation, followed by a period and the lemma itself.

8. Mathematical notation must be typeset. Equations should be numbered consecutively with Arabic numerals in parentheses placed flush right, and should be thusly referred to in the text [such as Eqs.(2) and (5)]. The running title must be placed at the top of even numbered pages and the first author's name, et al., must be placed at the top of the odd numbered pages.

9. Illustrations (photographs, drawings, diagrams, and charts) are to be numbered in one consecutive series of Arabic numerals. The captions for illustrations should be typed double space. All illustrations, charts, tables, etc., must be embedded in the body of the manuscript in proper, final, print position. In particular, manuscript, source, and PDF file version must be at camera ready stage for publication or they cannot be considered.

Tables are to be numbered (with Roman numerals) and referred to by number in the text. Center the title above the table, and type explanatory footnotes (indicated by superscript lowercase letters) below the table.

10. List references alphabetically at the end of the paper and number them consecutively. Each must be cited in the text by the appropriate Arabic numeral in square brackets on the baseline.

References should include (in the following order):  
initials of first and middle name, last name of author(s)  
title of article,

name of publication, volume number, inclusive pages, and year of publication.

Authors should follow these examples:

### **Journal Article**

1. H.H.Gonska, Degree of simultaneous approximation of bivariate functions by Gordon operators, (journal name in italics) *J. Approx. Theory*, 62,170-191(1990).

### **Book**

2. G.G.Lorentz, (title of book in italics) *Bernstein Polynomials* (2nd ed.), Chelsea, New York, 1986.

### **Contribution to a Book**

3. M.K.Khan, Approximation properties of beta operators, in (title of book in italics) *Progress in Approximation Theory* (P.Nevai and A.Pinkus, eds.), Academic Press, New York, 1991, pp.483-495.

11. All acknowledgements (including those for a grant and financial support) should occur in one paragraph that directly precedes the References section.

12. Footnotes should be avoided. When their use is absolutely necessary, footnotes should be numbered consecutively using Arabic numerals and should be typed at the bottom of the page to which they refer. Place a line above the footnote, so that it is set off from the text. Use the appropriate superscript numeral for citation in the text.

13. After each revision is made please again submit three hard copies of the revised manuscript, including in the final one. And after a manuscript has been accepted for publication and with all revisions incorporated, manuscripts, including the TEX/LaTeX source file and the PDF file, are to be submitted to the Editor's Office on a personal-computer disk, 3.5 inch size. Label the disk with clearly written identifying information and properly ship, such as:

Your name, title of article, kind of computer used, kind of software and version number, disk format and files names of article, as well as abbreviated journal name.

Package the disk in a disk mailer or protective cardboard. Make sure contents of disks are identical with the ones of final hard copies submitted!

Note: The Editor's Office cannot accept the disk without the accompanying matching hard copies of manuscript. No e-mail final submissions are allowed! The disk submission must be used.

14. Effective 1 Nov. 2009 for current journal page charges, contact the Editor in Chief. Upon acceptance of the paper an invoice will be sent to the contact author. The fee payment will be due one month from the invoice date. The article will proceed to publication only after the fee is paid. The charges are to be sent, by money order or certified check, in US dollars, payable to Eudoxus Press, LLC, to the address shown on

the Eudoxus [homepage](#).

No galleys will be sent and the contact author will receive one(1) electronic copy of the journal issue in which the article appears.

15. This journal will consider for publication only papers that contain proofs for their listed results.

# **TABLE OF CONTENTS, JOURNAL OF CONCRETE AND APPLICABLE MATHEMATICS, VOL. 8, NO. 2, 2010**

<b>Approximation by Nonlinear Bernstein and Favard-Szasz-Mirakjan Operators of Max-Product Kind, Barnabas Bede, Sorin G. Gal,.....</b>	<b>193</b>
<b>L-approximation to non-periodic functions, Michael I. Ganzburg, .....</b>	<b>208</b>
<b>Inequalities for Self-Reciprocal Polynomials and Uniformly Almost Periodic Functions, N. K. Govil, Q.M.Tariq,.....</b>	<b>216</b>
<b>Numerical approximation to pi using parabolic segments, Mark Bollman, George Grossman,.....</b>	<b>236</b>
<b>On the uniform spectrum of bounded functions and applications to differential equations, Nguyen van Minh, Gisele M. Mophou, Gaston N'Guerekata,.....</b>	<b>246</b>
<b>Sequential Decision Model for Exponential Pattern Recognition, Iuliana Florentina Iatan,.....</b>	<b>261</b>
<b>Inequalities for polynomials with curved majorants dependent on Chebyshev polynomials, Sergey Ivanovich Kalmykov,.....</b>	<b>270</b>
<b>Polygonal approximations of non-rectifiable curves and the jump problem, B.A. Kats,...</b>	<b>284</b>
<b>The boundary value problem for the fourth-order equation with fractional derivatives, D. Amanov, N.M. Kuzibaev,.....</b>	<b>292</b>
<b>On Sparse Solutions of Underdetermined Linear Systems, Ming-Jun Lai,.....</b>	<b>296</b>
<b>Patterns P, Milton del Castillo Lesmes Acosta,.....</b>	<b>328</b>
<b>A note on the construction of the <math>sl(2, \mathbb{R})</math> integral for ordinary differential equations of maximal symmetry, Sibusiso Moyo,.....</b>	<b>336</b>
<b>MAC solution for a rectangular membrane, Igor Neygebauer,.....</b>	<b>344</b>
<b>Multiresolution Inverse Wavelet Reconstruction from a Fourier Partial Sum, Nataniel Greene,.....</b>	<b>353</b>
<b>The calculation of axisymmetric duct geometries for incompressible, rotational flow using an integral formula based on Green's Theorem, Vasos Pavlika,.....</b>	<b>371</b>

**VOLUME 8, NUMBER 3      JULY 2010**

**ISSN:1548-5390 PRINT,1559-176X ONLINE**



**JOURNAL  
OF CONCRETE  
AND APPLICABLE**

**MATHEMATICS  
SPECIAL ISSUE III :APPLIED MATHEMATICS  
AND APPROXIMATION THEORY**

**EUDOXUS PRESS,LLC**

**SCOPE AND PRICES OF THE JOURNAL**  
**Journal of Concrete and Applicable Mathematics**

A quartely international publication of **Eudoxus Press,LLC**

**Editor in Chief: George Anastassiou**

Department of Mathematical Sciences,  
 University of Memphis  
 Memphis, TN 38152, U.S.A.  
 ganastss@memphis.edu

The main purpose of the "Journal of Concrete and Applicable Mathematics" is to publish high quality original research articles from all subareas of Non-Pure and/or Applicable Mathematics and its many real life applications, as well connections to other areas of Mathematical Sciences, as long as they are presented in a Concrete way. It welcomes also related research survey articles and book reviews. A sample list of connected mathematical areas with this publication includes and is not restricted to: Applied Analysis, Applied Functional Analysis, Probability theory, Stochastic Processes, Approximation Theory, O.D.E, P.D.E, Wavelet, Neural Networks, Difference Equations, Summability, Fractals, Special Functions, Splines, Asymptotic Analysis, Fractional Analysis, Inequalities, Moment Theory, Numerical Functional Analysis, Tomography, Asymptotic Expansions, Fourier Analysis, Applied Harmonic Analysis, Integral Equations, Signal Analysis, Numerical Analysis, Optimization, Operations Research, Linear Programming, Fuzzyness, Mathematical Finance, Stochastic Analysis, Game Theory, Math. Physics aspects, Applied Real and Complex Analysis, Computational Number Theory, Graph Theory, Combinatorics, Computer Science Math. related topics, combinations of the above, etc. In general any kind of Concretely presented Mathematics which is Applicable fits to the scope of this journal. Working Concretely and in Applicable Mathematics has become a main trend in many recent years, so we can understand better and deeper and solve the important problems of our real and scientific world. "Journal of Concrete and Applicable Mathematics" is a peer-reviewed International Quarterly Journal. We are calling for papers for possible publication. The contributor should send three copies of the contribution to the editor in-Chief typed in TEX, LATEX double spaced. [ See: Instructions to Contributors]

**Journal of Concrete and Applicable Mathematics(JCAAM)**

**ISSN:1548-5390 PRINT, 1559-176X ONLINE.**

is published in January, April, July and October of each year by

**EUDOXUS PRESS,LLC,**

1424 Beaver Trail Drive, Cordova, TN38016, USA,

Tel.001-901-751-3553

anastassioug@yahoo.com

<http://www.EudoxusPress.com>.

**Visit also [www.msci.memphis.edu/~ganastss/jcaam](http://www.msci.memphis.edu/~ganastss/jcaam).**

**Webmaster: Ray Clapsadle**

**Annual Subscription Current Prices:** For USA and Canada, Institutional: Print \$400, Electronic \$250, Print and Electronic \$450. Individual: Print \$150, Electronic

\$80,Print &Electronic \$200.For any other part of the world add \$50 more to the above prices for Print.

Single article PDF file for individual \$15.Single issue in PDF form for individual \$60.

No credit card payments.Only certified check,money order or international check in US dollars are acceptable.

Combination orders of any two from JoCAAA,JCAAM,JAFa receive 25% discount,all three receive 30% discount.

**Copyright**©2010 by Eudoxus Press,LLC all rights reserved.JCAAM is printed in USA.

**JCAAM is reviewed and abstracted by AMS Mathematical Reviews,MATHSCI,and Zentralblatt MATH.**

It is strictly prohibited the reproduction and transmission of any part of JCAAM and in any form and by any means without the written permission of the publisher.It is only allowed to educators to Xerox articles for educational purposes.The publisher assumes no responsibility for the content of published papers.

***JCAAM IS A JOURNAL OF RAPID PUBLICATION***

---

## Editorial Board

### Associate Editors

---

**Editor in -Chief:**

George Anastassiou  
 Department of Mathematical Sciences  
 The University Of Memphis  
 Memphis, TN 38152, USA  
 tel. 901-678-3144, fax 901-678-2480  
 e-mail ganastss@memphis.edu  
[www.msci.memphis.edu/~anastasg/anlyjour.htm](http://www.msci.memphis.edu/~anastasg/anlyjour.htm)  
 Areas: Approximation Theory,  
 Probability, Moments, Wavelet,  
 Neural Networks, Inequalities, Fuzzyness.

**Associate Editors:**

1) Ravi Agarwal  
 Florida Institute of Technology  
 Applied Mathematics Program  
 150 W. University Blvd.  
 Melbourne, FL 32901, USA  
[agarwal@fit.edu](mailto:agarwal@fit.edu)  
 Differential Equations, Difference  
 Equations,  
 Inequalities

2) Drumi D. Bainov  
 Medical University of Sofia  
 P.O. Box 45, 1504 Sofia, Bulgaria  
[drumibainov@yahoo.com](mailto:drumibainov@yahoo.com)  
 Differential Equations, Optimal Control,  
 Numerical Analysis, Approximation Theory

3) Carlo Bardaro  
 Dipartimento di Matematica & Informatica  
 Università di Perugia  
 Via Vanvitelli 1  
 06123 Perugia, ITALY  
 tel. +390755855034, +390755853822,  
 fax +390755855024  
[bardaro@unipg.it](mailto:bardaro@unipg.it) ,  
[bardaro@dipmat.unipg.it](mailto:bardaro@dipmat.unipg.it)  
 Functional Analysis and Approximation Th.,  
 Summability, Signal Analysis, Integral  
 Equations,  
 Measure Th., Real Analysis

4) Francoise Bastin  
 Institute of Mathematics  
 University of Liege  
 4000 Liege

21) Gustavo Alberto Perla Menzala  
 National Laboratory of Scientific Computation  
 LNCC/MCT  
 Av. Getulio Vargas 333  
 25651-075 Petropolis, RJ  
 Caixa Postal 95113, Brasil  
 and

Federal University of Rio de Janeiro  
 Institute of Mathematics  
 RJ, P.O. Box 68530 Rio de Janeiro, Brasil  
[perla@lncc.br](mailto:perla@lncc.br) and [perla@im.ufrj.br](mailto:perla@im.ufrj.br)  
 Phone 55-24-22336068, 55-21-25627513 Ext 224  
 FAX 55-24-22315595  
 Hyperbolic and Parabolic Partial Differential  
 Equations,  
 Exact controllability, Nonlinear Lattices and  
 Global  
 Attractors, Smart Materials

22) Ram N. Mohapatra  
 Department of Mathematics  
 University of Central Florida  
 Orlando, FL 32816-1364  
 tel. 407-823-5080  
[ramm@pegasus.cc.ucf.edu](mailto:ramm@pegasus.cc.ucf.edu)  
 Real and Complex analysis, Approximation Th.,  
 Fourier Analysis, Fuzzy Sets and Systems

23) Rainer Nagel  
 Arbeitsbereich Funktionalanalysis  
 Mathematisches Institut  
 Auf der Morgenstelle 10  
 D-72076 Tuebingen  
 Germany  
 tel. 49-7071-2973242  
 fax 49-7071-294322  
[rana@fa.uni-tuebingen.de](mailto:rana@fa.uni-tuebingen.de)  
 evolution equations, semigroups, spectral th.,  
 positivity

24) Panos M. Pardalos  
 Center for Appl. Optimization  
 University of Florida  
 303 Weil Hall  
 P.O. Box 116595  
 Gainesville, FL 32611-6595  
 tel. 352-392-9011  
[pardalos@ufl.edu](mailto:pardalos@ufl.edu)  
 Optimization, Operations Research



## BELGIUM

f.bastin@ulg.ac.be  
Functional Analysis, Wavelets

5) Yeol Je Cho  
Department of Mathematics Education  
College of Education  
Gyeongsang National University  
Chinju 660-701

## KOREA

tel. 055-751-5673 Office,  
055-755-3644 home,  
fax 055-751-6117  
yjcho@nongae.gsnu.ac.kr  
Nonlinear operator Th., Inequalities,  
Geometry of Banach Spaces

6) Sever S. Dragomir  
School of Communications and Informatics  
Victoria University of Technology  
PO Box 14428  
Melbourne City M.C  
Victoria 8001, Australia  
tel 61 3 9688 4437, fax 61 3 9688 4050  
sever.dragomir@vu.edu.au,  
sever@sci.vu.edu.au  
Math. Analysis, Inequalities, Approximation  
Th.,  
Numerical Analysis, Geometry of Banach  
Spaces,  
Information Th. and Coding

7) Angelo Favini  
Università di Bologna  
Dipartimento di Matematica  
Piazza di Porta San Donato 5  
40126 Bologna, ITALY  
tel. ++39 051 2094451  
fax. ++39 051 2094490  
favini@dm.unibo.it  
Partial Differential Equations, Control  
Theory,  
Differential Equations in Banach Spaces

8) Claudio A. Fernandez  
Facultad de Matematicas  
Pontificia Universidad Católica de Chile  
Vicuna Mackenna 4860  
Santiago, Chile  
tel. ++56 2 354 5922  
fax. ++56 2 552 5916  
cfernand@mat.puc.cl  
Partial Differential Equations,  
Mathematical Physics,  
Scattering and Spectral Theory

25) Svetlozar T. Rachev  
Dept. of Statistics and Applied Probability  
Program

University of California, Santa Barbara  
CA 93106-3110, USA  
tel. 805-893-4869  
rachev@pstat.ucsb.edu

## AND

Chair of Econometrics and Statistics  
School of Economics and Business Engineering  
University of Karlsruhe  
Kollegium am Schloss, Bau II, 20.12, R210  
Postfach 6980, D-76128, Karlsruhe, Germany  
tel. 011-49-721-608-7535  
rachev@lsoe.uni-karlsruhe.de  
Mathematical and Empirical Finance,  
Applied Probability, Statistics and Econometrics

26) John Michael Rassias  
University of Athens  
Pedagogical Department  
Section of Mathematics and Informatics  
20, Hippocratous Str., Athens, 106 80, Greece

Address for Correspondence

4, Agamemnonos Str.  
Aghia Paraskevi, Athens, Attikis 15342 Greece  
jrassias@primedu.uoa.gr  
jrassias@tellas.gr  
Approximation Theory, Functional Equations,  
Inequalities, PDE

27) Paolo Emilio Ricci  
Università degli Studi di Roma "La Sapienza"  
Dipartimento di Matematica-Istituto  
"G. Castelnuovo"  
P.le A. Moro, 2-00185 Roma, ITALY  
tel. ++39 0649913201, fax ++39 0644701007  
riccip@uniroma1.it, Paoloemilio.Ricci@uniroma1.it  
Orthogonal Polynomials and Special functions,  
Numerical Analysis, Transforms, Operational  
Calculus,  
Differential and Difference equations

28) Cecil C. Rousseau  
Department of Mathematical Sciences  
The University of Memphis  
Memphis, TN 38152, USA  
tel. 901-678-2490, fax 901-678-2480  
ccrousse@memphis.edu  
Combinatorics, Graph Th.,  
Asymptotic Approximations,  
Applications to Physics

29) Tomasz Rychlik

## 9) A.M.Fink

Department of Mathematics  
Iowa State University  
Ames, IA 50011-0001, USA  
tel. 515-294-8150  
fink@math.iastate.edu  
Inequalities, Ordinary Differential  
Equations

## 10) Sorin Gal

Department of Mathematics  
University of Oradea  
Str. Armatei Romane 5  
3700 Oradea, Romania  
galso@uoradea.ro  
Approximation Th., Fuzzyness, Complex  
Analysis

## 11) Jerome A. Goldstein

Department of Mathematical Sciences  
The University of Memphis,  
Memphis, TN 38152, USA  
tel. 901-678-2484  
jgoldste@memphis.edu  
Partial Differential Equations,  
Semigroups of Operators

## 12) Heiner H. Gonska

Department of Mathematics  
University of Duisburg  
Duisburg, D-47048  
Germany  
tel. 0049-203-379-3542 office  
gonska@informatik.uni-duisburg.de  
Approximation Th., Computer Aided  
Geometric Design

## 13) Dmitry Khavinson

Department of Mathematical Sciences  
University of Arkansas  
Fayetteville, AR 72701, USA  
tel. (479) 575-6331, fax (479) 575-8630  
dmitry@uark.edu  
Potential Th., Complex Analysis, Holomorphic  
PDE,  
Approximation Th., Function Th.

## 14) Virginia S. Kiryakova

Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
virginia@diogenes.bg  
Special Functions, Integral Transforms,  
Fractional Calculus

## 15) Hans-Bernd Knoop

Institute of Mathematics  
Polish Academy of Sciences  
Chopina 12, 87100 Torun, Poland  
T.Rychlik@impan.gov.pl  
Mathematical Statistics, Probabilistic  
Inequalities

## 30) Bl. Sendov

Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
bsendov@bas.bg  
Approximation Th., Geometry of Polynomials,  
Image Compression

## 31) Igor Shevchuk

Faculty of Mathematics and Mechanics  
National Taras Shevchenko  
University of Kyiv  
252017 Kyiv  
UKRAINE  
shevchuk@univ.kiev.ua  
Approximation Theory

## 32) H.M. Srivastava

Department of Mathematics and Statistics  
University of Victoria  
Victoria, British Columbia V8W 3P4  
Canada  
tel. 250-721-7455 office, 250-477-6960 home,  
fax 250-721-8962  
harimsri@math.uvic.ca  
Real and Complex Analysis, Fractional Calculus  
and Appl.,  
Integral Equations and Transforms, Higher  
Transcendental  
Functions and Appl., q-Series and q-Polynomials,  
Analytic Number Th.

## 33) Stevo Stevic

Mathematical Institute of the Serbian Acad. of  
Science  
Knez Mihailova 35/I  
11000 Beograd, Serbia  
sstevic@ptt.yu; sstevo@matf.bg.ac.yu  
Complex Variables, Difference Equations,  
Approximation Th., Inequalities

## 34) Ferenc Szidarovszky

Dept. Systems and Industrial Engineering  
The University of Arizona  
Engineering Building, 111  
PO. Box 210020  
Tucson, AZ 85721-0020, USA  
szidar@sie.arizona.edu  
Numerical Methods, Game Th., Dynamic Systems,

Institute of Mathematics  
 Gerhard Mercator University  
 D-47048 Duisburg  
 Germany  
 tel.0049-203-379-2676  
 knoop@math.uni-duisburg.de  
 Approximation Theory, Interpolation

16) Jerry Koliha  
 Dept. of Mathematics & Statistics  
 University of Melbourne  
 VIC 3010, Melbourne  
 Australia  
 koliha@unimelb.edu.au  
 Inequalities, Operator Theory,  
 Matrix Analysis, Generalized Inverses

17) Mustafa Kulenovic  
 Department of Mathematics  
 University of Rhode Island  
 Kingston, RI 02881, USA  
 kulenm@math.uri.edu  
 Differential and Difference Equations

18) Gerassimos Ladas  
 Department of Mathematics  
 University of Rhode Island  
 Kingston, RI 02881, USA  
 gladas@math.uri.edu  
 Differential and Difference Equations

19) V. Lakshmikantham  
 Department of Mathematical Sciences  
 Florida Institute of Technology  
 Melbourne, FL 32901  
 e-mail: lakshmik@fit.edu  
 Ordinary and Partial Differential  
 Equations,  
 Hybrid Systems, Nonlinear Analysis

20) Rupert Lasser  
 Institut für Biomathematik & Biomertie, GSF  
 -National Research Center for environment  
 and health  
 Ingolstaedter landstr.1  
 D-85764 Neuherberg, Germany  
 lasser@gsf.de  
 Orthogonal Polynomials, Fourier Analysis,  
 Mathematical Biology

Multicriteria Decision making,  
 Conflict Resolution, Applications  
 in Economics and Natural Resources  
 Management

35) Gancho Tachev  
 Dept. of Mathematics  
 Univ. of Architecture, Civil Eng. and Geodesy  
 1 Hr. Smirnenski blvd  
 BG-1421 Sofia, Bulgaria  
 gtt\_fte@uacg.bg  
 Approximation Theory

36) Manfred Tasche  
 Department of Mathematics  
 University of Rostock  
 D-18051 Rostock  
 Germany  
 manfred.tasche@mathematik.uni-rostock.de  
 Approximation Th., Wavelet, Fourier Analysis,  
 Numerical Methods, Signal Processing,  
 Image Processing, Harmonic Analysis

37) Chris P. Tsokos  
 Department of Mathematics  
 University of South Florida  
 4202 E. Fowler Ave., PHY 114  
 Tampa, FL 33620-5700, USA  
 profcpt@math.usf.edu, profcpt@chumal.cas.usf.edu  
 Stochastic Systems, Biomathematics,  
 Environmental Systems, Reliability Th.

38) Lutz Volkmann  
 Lehrstuhl II für Mathematik  
 RWTH-Aachen  
 Templergraben 55  
 D-52062 Aachen  
 Germany  
 volkm@math2.rwth-aachen.de  
 Complex Analysis, Combinatorics, Graph Theory

### **EDITOR'S NOTE**

This special issue on “Applied Mathematics and Approximation Theory” contains expanded versions of articles that were presented in the international conference “Applied Mathematics and Approximation Theory 2008” ( AMAT 08), during October 11-13, 2008 at the University of Memphis, Memphis, Tennessee, USA. All articles were refereed.

The organizer and Editor

George Anastassiou

# Algorithms for Segmentwise Computation of Forward and Inverse Discrete-time Wavelet Transform

Pavel Rajmic

Faculty of Electrical Engineering and Communications

Brno University of Technology

Purkyňova 118, 612 00, Brno, Czech Republic

E-mail: rajmic@feec.vutbr.cz

## Abstract

The paper describes a method of segmented wavelet transform (SegWT) that makes it possible to compute the discrete-time wavelet transform of a signal segment-by-segment, with exactly the same result as if the whole signal were transformed at once. Due to its generality, the method can be utilized in many situations: for wavelet-type processing of a signal in real time or in case we want to process the signal in parallel or in case we need to process a long signal, but the available memory capacity is insufficient (e.g. in the DSPs). In the paper, the background theory and the emerging principles of both the forward and the inverse SegWT are explained.

**Keywords** segmented discrete-time wavelet transform, real-time wavelet transform, SegWT, signal processing, algorithm

## 1 Introduction

The discrete-time wavelet transform (DTWT) has many applications in the field of signal and image processing today. Most of them require signals to be completely known when the processing is initiated. The natural need for real-time applications originated in the development of methods allowing the computation of the DTWT without knowing the signal in advance, and possibly with minimum delay at the same time. Applications arising in the image processing field such as wavelet compression segment-by-segment (e.g. JPEG2000 coding for large images) lead to the same category of methods.

We are not interested in algorithms for transforming data received as a continuous stream [4] (e.g. in the FPGAs), because we suppose the data come in the form of nonoverlapping segments. Leaving these algorithms aside, methods



Figure 1: The undesirable artefacts near the segments borders. The image was (strongly) compressed using JPEG2000 algorithm, with tiling option switched on; the square tiles of  $128 \times 128$  px in size — segments — are transformed separately.

for computing the wavelet transform in succession can be divided into two main classes.

### 1.1 Inexact Methods

This class includes another two types of methods:

The first-type methods are based on signal windowing and overlapping the resultant segments [2], analogous to the short-time Fourier analysis. However, windowing introduces (apart from potential considerable numerical errors at the window tails) severe problems when the wavelet coefficients are subject to nonlinear processing, e.g. the thresholding step within a denoising algorithm.

The second-type methods approach the particular signal segments independently — they “blindly” extrapolate samples beyond the boundaries of each segment. There exist several such methods, either simpler or more complex [7]. This approach, of course, leads to undesirable artifacts on the signal boundaries after the processing.

A typical example of the described inexactness can be seen in Fig. 1.

### 1.2 Exact Methods

This class, which we are most interested in, includes methods which actually extend the boundaries by samples from the respective neighbouring segments. Most of the contributions in these problems have come from researchers working with images and video. They were motivated by the idea of splitting the com-

putation into parallel processes run on individual workstations [3, 1]. Within this class of algorithms, we can distinguish another two algorithm types:

The third-type algorithms split the signal into segments which are distributed to the particular processes. During the whole computation the processes do not communicate with each other. Afterwards, the results are joined together. Such an approach is suitable for systems where interchanging information (i.e. wavelet coefficients) is slow and thus increases the total computation time. The disadvantage is that a portion of computation is performed redundantly, in several processes.

In the algorithms of the fourth type the processes mutually interchange data during computation. Clearly, this is suitable in situations where the communication is fast. The principal advantage here is that there is no computation redundancy. To be more concrete, wavelet coefficients located by the segment boundaries computed at each “level” of the transformation are interchanged.

State-of-the-art algorithms belonging to both the third and the fourth group, which can be found in the literature, are derived for the special case when each segment length is equal to a power of two. This assumption is their drawback, mainly for larger segments (e.g. the difference between 1024 and 2048 can be inadmissibly big, considering for example that with images,  $1024^2 \doteq 10^6$  and  $2048^2 \doteq 4 \cdot 10^6$ ). Also, there are situations where the segment sizes are not a power of two (e.g. the signal buffer size in audio cards running with ASIO driver could be 96 samples). Paper [5] gives directions for the non-power-of-two case. However, there is one more thing: all of the algorithms mentioned are made just for the purpose of forward DTWT, as they are mainly used for blockwise image compression. Moreover, most of the published methods specialize in JPEG2000, which means that they are restricted to the biorthogonal wavelet CDF 9/7.

### 1.3 Motivation and Goal of SegWT

The objective for the segmented wavelet transform, denoted SegWT, is naturally to allow signal processing by its segments, so that in this manner we get the same result (i.e. the same wavelet coefficients) as in the common DTWT case. At the same time, SegWT should utilize as much from the DTWT computational routines as possible.

In this paper we present the SegWT method, which can be utilized for any wavelet-type segmentwise data processing task, that is to say also in real time.

SegWT includes both the forward and the inverse parts of the transform. The segment length and the wavelet filter can be chosen arbitrarily. In fact, SegWT is of the third type in the sense of the above.

The derivation of the SegWT algorithm requires a very detailed knowledge of the behavior of ordinary DTWT, so before we start with SegWT, we recall the basic algorithm of DTWT.

## 2 Classical DTWT Algorithm

**Algorithm 1** (*Decomposition pyramidal algorithm DTWT*) Let  $\mathbf{x}$  be a discrete input signal of length  $s$ , the two wavelet decomposition filters of length  $m$  are defined, highpass  $\mathbf{g}$  and lowpass  $\mathbf{h}$ ,  $J$  is a positive integer denoting the decomposition depth. Also, the type of boundary treatment [7, ch. 8] has to be known.

1. Denote the input signal  $\mathbf{x}$  by  $\mathbf{a}^{(0)}$  and set  $j = 0$ .
2. One decomposition step:
  - (a) Extending the input vector. Extend  $\mathbf{a}^{(j)}$  from both the left and the right sides by  $(m - 1)$  samples, according to the type of boundary treatment.
  - (b) Filtering. Convolve the extended signal with filter  $\mathbf{g}$ .
  - (c) Cropping. Take from the result just its central part, so that the remaining “tails” on both the left and the right sides have the same length  $m - 1$  samples.
  - (d) Decimation. Downsample the resultant vector.

Denote the resulting vector by  $\mathbf{d}^{(j+1)}$  and store it. Repeat items b)–d), now with filter  $\mathbf{h}$ , denoting and storing the result as  $\mathbf{a}^{(j+1)}$ .

3. Increase  $j$  by one. If it now holds  $j < J$ , return to item 2, in the other cases the algorithm ends.

After Algorithm 1 has been finished, we have the wavelet coefficients stored in  $J + 1$  vectors (of different length)  $\mathbf{a}^{(J)}, \mathbf{d}^{(J)}, \mathbf{d}^{(J-1)}, \dots, \mathbf{d}^{(1)}$ .

## 3 Method of Segmented Wavelet Transform

In the problem, the following parameters play a crucial role:  $m \dots$  wavelet filter length,  $m > 0$ ,  $J \dots$  transform depth,  $J > 0$ ,  $s \dots$  length of the segment,  $s > 0$ .

Based on a detailed knowledge of DTWT, it is possible to deduce fairly sophisticated rules how to handle the signal segments. It is clear that it is necessary to extend every segment from the left by an exact number of samples from the preceding segment, and from the right by another number of samples from the subsequent segment (extension, overlap). However, the number of such samples depends on  $m, J$  and  $s$ , and it can be shown that every segment has to be extended by a different length from the left and from the right, and these lengths can also differ from segment to segment! And, of course, the first and the last segments have to be handled in a particular way.



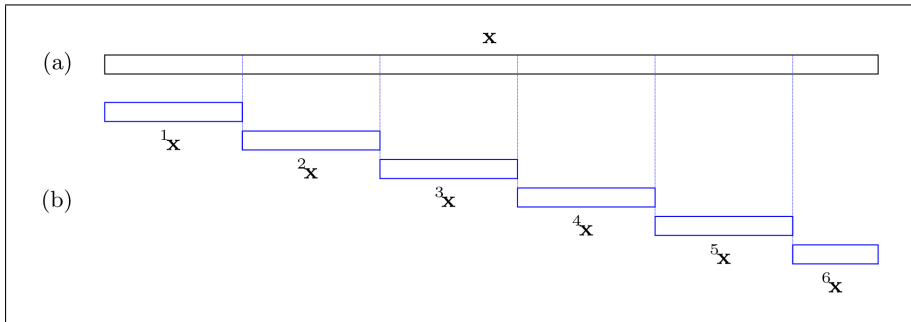


Figure 2: Scheme of signal segmentation. The input signal  $\mathbf{x}$  (a) is divided into segments of equal length and the last one can be shorter than this (b).

### 3.1 Important Theorems Derived from DTWT Algorithm

Before we introduce a detailed description of the SegWT algorithm, several theorems must be presented. More theorems including proofs can be found in [6, ch.8]. We assume that the input signal  $\mathbf{x}$  is divided into  $S \geq 1$  segments of equal length  $s$ . Single segments will be denoted  $^1\mathbf{x}, ^2\mathbf{x}, \dots, ^S\mathbf{x}$ . The last one can be less long than  $s$  and the number  $S$  does not have to be known in advance. See Fig. 2. The signal boundary treatment considered in this paper is “zero-padding”, when the boundaries are extended by zeros (most suitable for processing audio recordings, for example), but switching to another type of treatment is easy.

By the formulation that *two sets of coefficients from the  $k$ -th decomposition level follow-up on each other* we mean a situation when two consecutive segments are properly extended, see Figures 2 and 3, so that applying the DTWT of depth  $k$ , with step 2a) omitted (cf. Algorithm 3, page 9), separately to both the segments, let us say  $^n\mathbf{x}$  and  $^{n+1}\mathbf{x}$ , and joining the resultant coefficients together leads to the situation that the last coefficient computed from  $^n\mathbf{x}$  and the first coefficient computed from  $^{n+1}\mathbf{x}$  would be neighboring in case the signal is transformed by the ordinary DTWT.

Such a situation is desirable and the theorems below lead to proper handling of the consecutive segments.

**Theorem 1** *In case that the consecutive segments have*

$$r(k) = (2^k - 1)(m - 1) \quad (1)$$

*common input signal samples, the coefficients from the  $k$ -th decomposition level follow-up on each other.*

Thus, for a decomposition depth equal to  $J$  it is necessary to have  $r(J) = (2^J - 1)(m - 1)$  common samples in the two consecutive segments after they have been extended. This extension must be divided into the right extension of the first segment (of length  $R$ ) and the left extension of the following segment (of

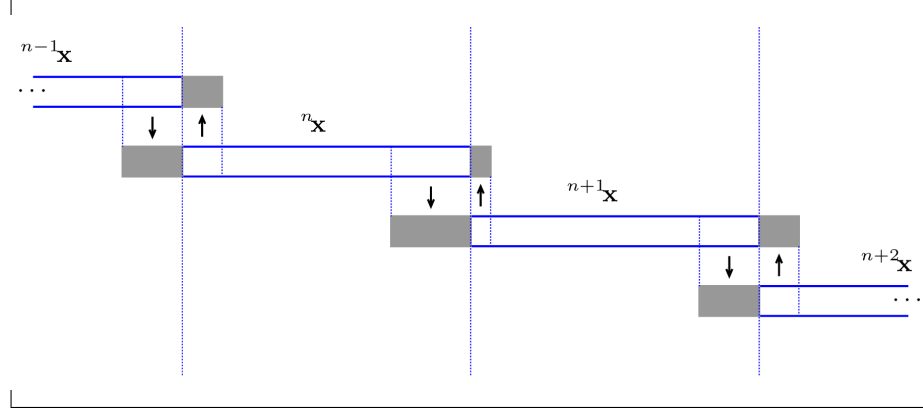


Figure 3: Illustration of extending the segments.

length  $L$ ), while  $r(J) = R + L$ . However, the lengths  $L, R \geq 0$  cannot be chosen arbitrarily. In general, the numbers  $L$  and  $R$  are not uniquely determined and must comply with strict rules that will be shown. The formula for the choice of extension  $L_{\max}$ , which is unique and the most appropriate in the case of real-time signal processing, is given in Theorem 2.

For the purpose of the following, we assign the number of the respective segment to the variables  $L_{\max}, R_{\min}, l$ , so that the left extension of the  $n$ -th segment will be of length  $L_{\max}(n)$ , the right extension will be of length  $R_{\min}(n)$  and the length of the original  $n$ -th segment with the left extension joined will be denoted  $l(n)$ .

**Theorem 2** *Let the  $n$ -th segment be given, whose length including its left extension is  $l(n)$ . The maximum possible left extension of the next segment,  $L_{\max}(n+1)$ , can be computed by the formula*

$$L_{\max}(n+1) = l(n) - 2^J \text{ceil} \left( \frac{l(n) - r(J)}{2^J} \right). \quad (2)$$

*The minimum possible right extension of the given segment is then*

$$R_{\min}(n) = r(J) - L_{\max}(n+1). \quad (3)$$

**Theorem 3** *The length of the right extension of the  $n$ -th segment, must comply with*

$$R_{\min}(n) = 2^J \text{ceil} \left( \frac{ns}{2^J} \right) - ns, \quad n = 1, 2, \dots, S-2. \quad (4)$$

From (4) it is clear that  $R_{\min}$  is periodic with respect to  $s$  with period  $2^J$ , i.e.  $R_{\min}(n+2^J) = R_{\min}(n)$ . This property, among other things, can be seen in Table I.

**Theorem 4** (on the total length of segment) *After the extension, the  $n$ -th segment of original length  $s$  will be of total length  $\sum(n)$ , which can acquire one of two values, either*

$$r(J) + 2^J \text{ceil}\left(\frac{s}{2^J}\right) \quad \text{or} \quad r(J) + 2^J \text{ceil}\left(\frac{s}{2^J}\right) - 2^J. \quad (5)$$

The overall illustration of SegWT can be seen in Fig.4. The particular algorithms are described in detail in the next sections.

### 3.2 Algorithm of Forward SegWT

The algorithm works such that it reads (receives) individual segments of the input signal, makes them extend each other in a proper way, then it computes the wavelet coefficients in a modified way and, in the end, it easily joins the coefficients. There is no need to know how many segments will be in total, we only require that in the moment when the last segment is received, we know that information.

**Algorithm 2** *Let the wavelet filters  $\mathbf{g}$  and  $\mathbf{h}$  be of length  $m$  and the decomposition depth be  $J$ . The boundary treatment mode is “zero-padding”. The segments of length  $s > 0$  of the input signal  $\mathbf{x}$  are denoted  ${}^1\mathbf{x}, {}^2\mathbf{x}, {}^3\mathbf{x}, \dots$ . The last segment contains  $s' \leq s$  samples.*

1. Set  $n = 1$ ,  $\text{last} = 0$ .
2. Read the first segment,  ${}^1\mathbf{x}$ . Extend it from the left by  $r(J)$  zero samples. Update ‘last’.
3. **If**, at the same time, the  $n$ -th segment is the last one
  - (a) Extend the  $n$ -th segment from the right by such a number of zero samples that its total length will be  $L_{\max}(n) + s$ .
  - (b) Extend the  $n$ -th segment from the right by  $r(J)$  zero samples.
  - (c) Compute the transform of depth  $J$  of the extended segment using Algorithm 3.
  - (d) Modify the vectors containing the wavelet coefficients by trimming off a certain number of redundant coefficients from the left side, specifically: on the  $k$ -th level,  $k = 1, 2, \dots, J - 1$ , trim off  $r(J - k)$  coefficients.
  - (e) Trim off redundant coefficients from the right so that on the  $k$ -th level floor  $(2^{-k}(L_{\max}(S) + s'))$  coefficients remain.
  - (f) Trim off the vectors in the same manner as in 3d, but this time from the right.
  - (g) Store the result as  ${}^n\mathbf{a}^{(J)}, {}^n\mathbf{d}^{(J)}, {}^n\mathbf{d}^{(J-1)}, \dots, {}^n\mathbf{d}^{(1)}$ .

**Otherwise**

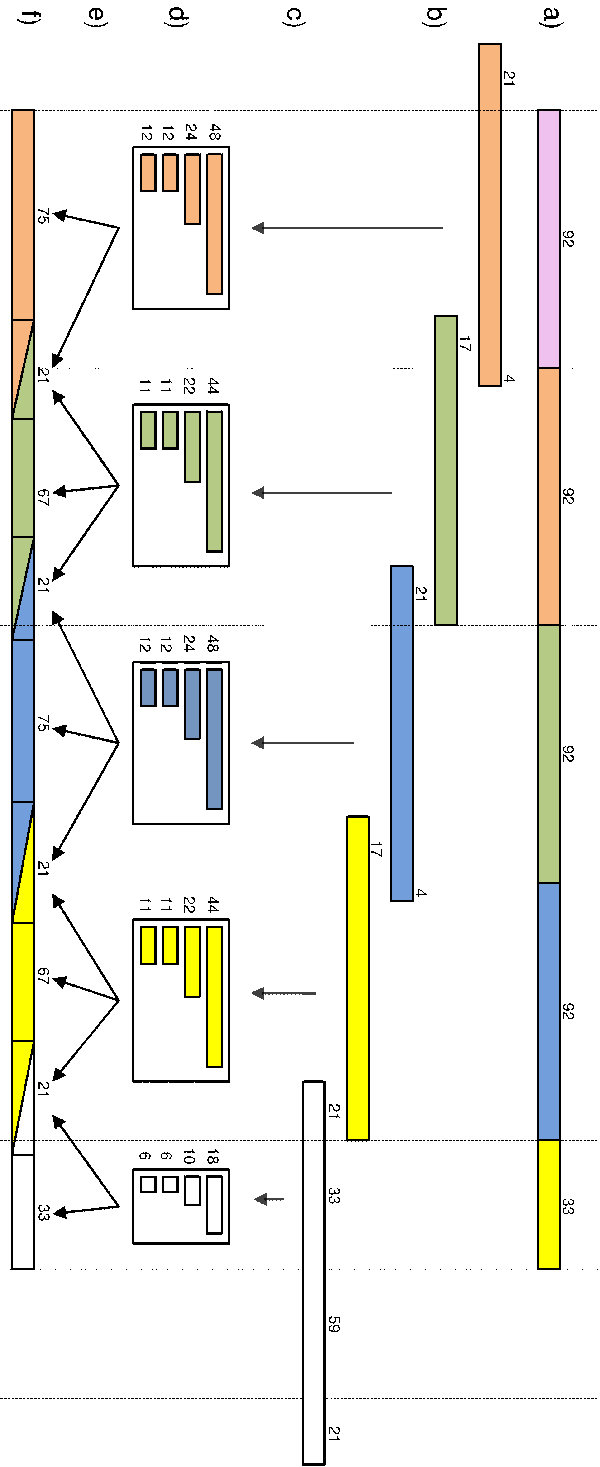


Figure 4: Overall scheme of the SegWT, forward and inverse parts, here in a particular case when  $s = 92$ : a) input signal segments, b) extending them (the left and right lengths differ from segment to segment), c) computation of the forward part, d) the computed blocks of coefficients, e) computation of the inverse part, f) the reconstructed signal.

- (h) Read the  $(n + 1)$ -th segment, update ‘last’.
- (i) Compute  $L_{\max}(n + 1)$  and  $R_{\min}(n)$  (see Theorem 2).
- (j) Extend the  $n$ -th segment from the right side:
  - If**  $\text{last} = 1$  (i.e. we have the last segment)
    - i. Compute the difference  $\text{diff} = \max(0, R_{\min}(n) - s')$ .
    - ii. **If**  $\text{diff} > 0$  (i.e. not enough samples in the last segment for extension by  $R_{\min}(n)$ )
      - A. Extend the  $n$ -th segment from the right side by  $s'$  samples from the last segment.
      - B. Extend the  $n$ -th segment from the right side by another  $\text{diff}$  zero samples.
    - Otherwise**
      - C. Extend the  $n$ -th segment from the right side by  $R_{\min}$  samples taken from the last segment.
  - iii. Extend the  $n$ -th segment from the right side by  $R_{\min}$  samples taken from the last segment.
- (k) Extend the  $(n+1)$ -th segment from the left side by  $L_{\max}(n+1)$  samples taken from segment  $n$ .
- (l) Compute the DTWT of depth  $J$  from the (extended)  $n$ -th segment using Algorithm 3.
- (m) Modify the particular vectors containing the coefficients in the same manner as in 3d.
- (n) Store the result as  ${}^n\mathbf{a}^{(J)}, {}^n\mathbf{d}^{(J)}, {}^n\mathbf{d}^{(J-1)}, \dots, {}^n\mathbf{d}^{(1)}$ .
- (o) Increase  $n$  by 1 and go to item 3.

**Algorithm 3** This sub-algorithm is identical to Algorithm 1 with the exception that we omit step 2a), i.e. we do not extend the vector.

The output of Algorithm 2 is  $S(J + 1)$  vectors of wavelet coefficients

$$\{ {}^i\mathbf{a}^{(J)}, {}^i\mathbf{d}^{(J)}, {}^i\mathbf{d}^{(J-1)}, \dots, {}^i\mathbf{d}^{(1)} \}_{i=1}^S. \quad (6)$$

If we simply join these vectors together, we obtain a set of  $J + 1$  vectors, which are identical to the wavelet coefficients of signal  $\mathbf{x}$ .

The flowchart of Algorithm 2 is in Fig. 5.

### 3.3 Corollaries and Limitations of SegWT Algorithm

In [6] several practical corollaries for SegWT can be found, e.g. that the segments cannot be shorter than  $2^J$ . From the description in the above sections it should be clear that the delay of Algorithm 2 is one segment (i.e.  $s$  samples) plus the time needed for the computation of the coefficient from the current segment. It can be easily shown that in the special case of  $s$  being divisible by  $2^J$  it even holds  $R_{\min}(n) = 0$  for every  $n \in \mathbb{N}$  (see Theorem 3), i.e. the delay of the forward method is determined only by the computation time!

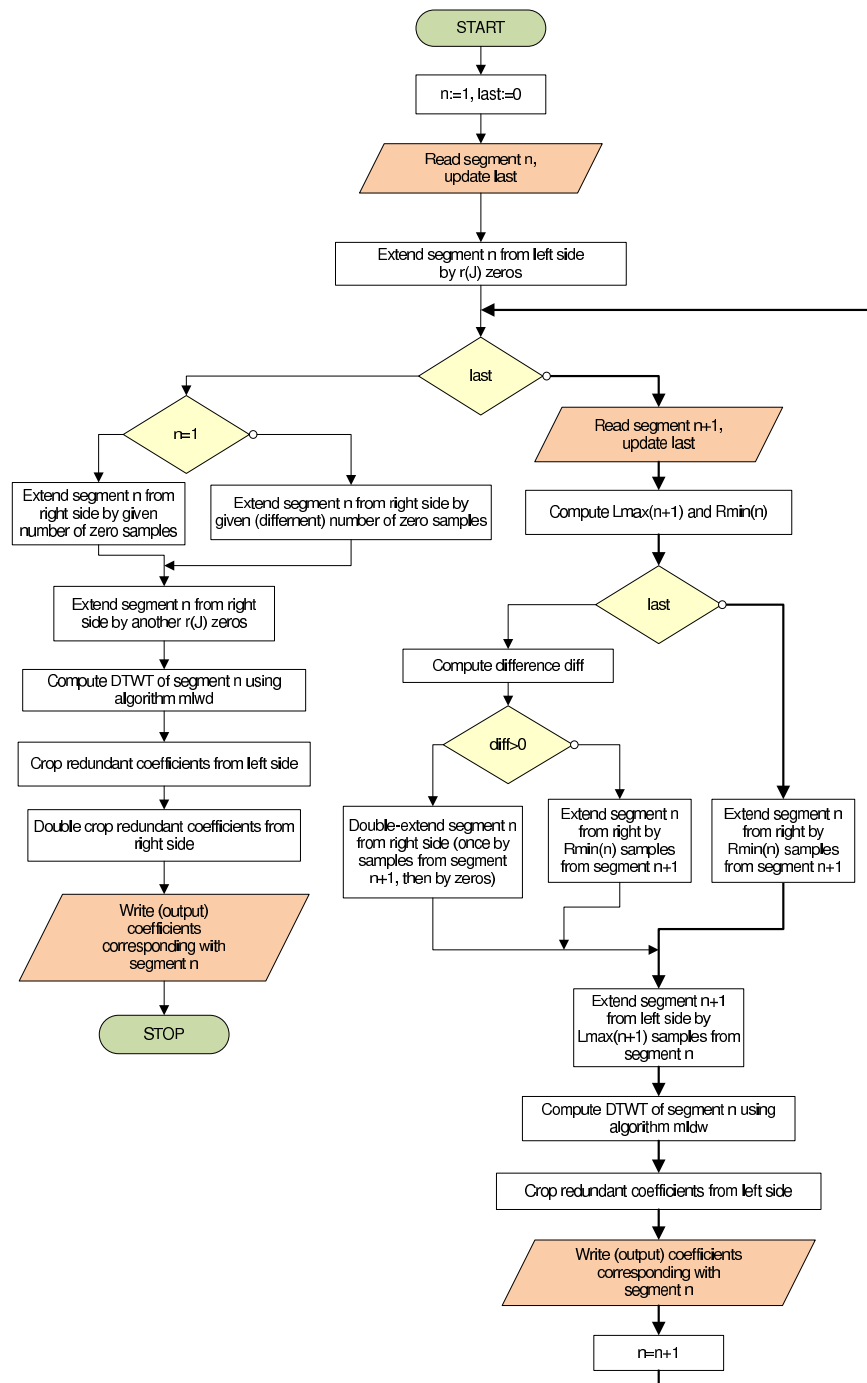


Figure 5: Flowchart of the forward SegWT, with zero-padding treatment of the signal boundaries. The main loop, which is applied to all the segment but the first and last ones, is emphasized by the thicker line.

Table I: Example — lengths of extensions for different lengths of segments  $s$ . The depth of decomposition is  $J = 3$  and the filter length is  $m = 16$ .

$s$	$n$	1	2	3	4	5	6	7	8	9	10	11	12	...
<b>512</b>	$L_{\max}(n)$	105	105	105	105	105	105	105	105	105	105	105	105	...
	$R_{\min}(n)$	0	0	0	0	0	0	0	0	0	0	0	0	...
	$\sum(n)$	617	617	617	617	617	617	617	617	617	617	617	617	...
<b>513</b>	$L_{\max}(n)$	105	98	99	100	101	102	103	104	105	98	99	100	...
	$R_{\min}(n)$	7	6	5	4	3	2	1	0	7	6	5	4	...
	$\sum(n)$	625	617	617	617	617	617	617	617	625	617	617	617	...
<b>514</b>	$L_{\max}(n)$	105	99	101	103	105	99	101	103	105	99	101	103	...
	$R_{\min}(n)$	6	4	2	0	6	4	2	0	6	4	2	0	...
	$\sum(n)$	625	617	617	617	625	617	617	617	625	617	617	617	...
<b>515</b>	$L_{\max}(n)$	105	100	103	98	101	104	99	102	105	100	103	98	...
	$R_{\min}(n)$	5	2	7	4	1	6	3	0	5	2	7	4	...
	$\sum(n)$	625	617	625	617	617	625	617	617	625	617	625	617	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$

### 3.4 Few Examples

- For  $J = 5$  and  $m = 8$ , the minimum segment length is 32 samples. When we set  $s = 256$ ,  $R_{\min}$  will always be zero and  $L_{\max} = r(5) = 217$ . The length of every extended segment will be  $256 + 217 = 473$  samples.
- For  $J = 5$  and  $m = 8$  we set  $s = 300$ , which is not divisible by  $2^5$ . Thus  $R_{\min}$  and  $L_{\max}$  will alternate with period 8 such that  $0 \leq R_{\min} \leq 31$  and  $186 \leq L_{\max} \leq 217$ . The total length of segment after extension will be either 505 or 537.
- (Example illustrated in Fig. 4) For  $J = 3$ ,  $m = 4$ ,  $s = 92$ , the extensions will alternate between two states, either  $R_{\min} = 4$  and  $L_{\max} = 17$  or  $R_{\min} = 0$  and  $L_{\max} = 21$ . The length of the extended segments will be 109 or 117 samples.

The increase of the samples entering the computation is naturally a price paid for the fact that no errors will originate during processing the boundaries.

### 3.5 Algorithm of Inverse SegWT

The inverse algorithm is described below, in less detail than the forward one. Blocks of wavelet coefficients (6) produced segment-by-segment by the forward SegWT constitute the input for the inverse algorithm. Analog to the forward case, we use the boolean flag *last*, which becomes true if the very last segment has to be processed.

In addition to that, due to the downsampling step of the forward transform, we loss information about the total length of the signal, more precisely we do not know if the original length was  $a$  or  $a + 1$  for some integer  $a$ . We could solve this problem by accumulating the lengths of individual inverted

segments, however, such a number could be very large, possibly overflowing the processor arithmetics. A better solution is just to keep the signal parity (i.e. if the accumulated length is even or odd). The information is then used at the very end of the signal for deciding to cut or not to cut the last reconstructed sample.

The inverse SegWT partly utilizes the overlap-add principle for joining the reconstructed pieces of the time-domain signal. The length of the overlap stays  $r(J)$  all the time. As for the illustration, we again refer to Fig. 4.

**Algorithm 4** *Let the decomposition depth  $J$  be given, as well as wavelet reconstruction filters  $\tilde{\mathbf{g}}$  and  $\tilde{\mathbf{h}}$  of length  $m$ , and coefficients  ${}^n\mathbf{a}^{(J)}, {}^n\mathbf{d}^{(J)}, {}^n\mathbf{d}^{(J-1)}, \dots, {}^n\mathbf{d}^{(1)}$  for all  $n$ .*

1. Set  $n = 1$ . Set  $\text{last} = 0$ .
2. **If**  $\text{last} = 1$  then the Algorithm ends.
3. Read the  $n$ -th block of coefficients and update 'last'.
4. Extend the detail coefficients: on the  $k$ -th level,  $k = 1, \dots, J - 1$ , append  $r(J - k)$  zero coefficients from the left side.
5. Compute the inverse transform of depth  $J$  using Algorithm 5.
6. **If**  $n \neq 1$ , recall the samples for the overlap, saved in the last cycle, and add them to the current inverted block.
7. Update the parity of the signal.
8. **If**  $\text{last} \neq 1$ , append the central, non-overlapping part to the output. Save the samples of the overlap of the current inverted segment for the next cycle.  
**Otherwise** Append the whole inversion to the output. Eventually crop several samples from the end of the signal.
9. The output (a segment of a time-domain signal) is now complete and prepared to be "sent".
10. Increase  $n$  by 1 and return to item 2.

**Algorithm 5** *This algorithm is identical to the ordinary inverse wavelet transform (i.e. upsampling – filtering – summing – cropping), but the cropping phase is omitted here.*

The flowchart of Algorithm 4 can be seen in Fig. 6.



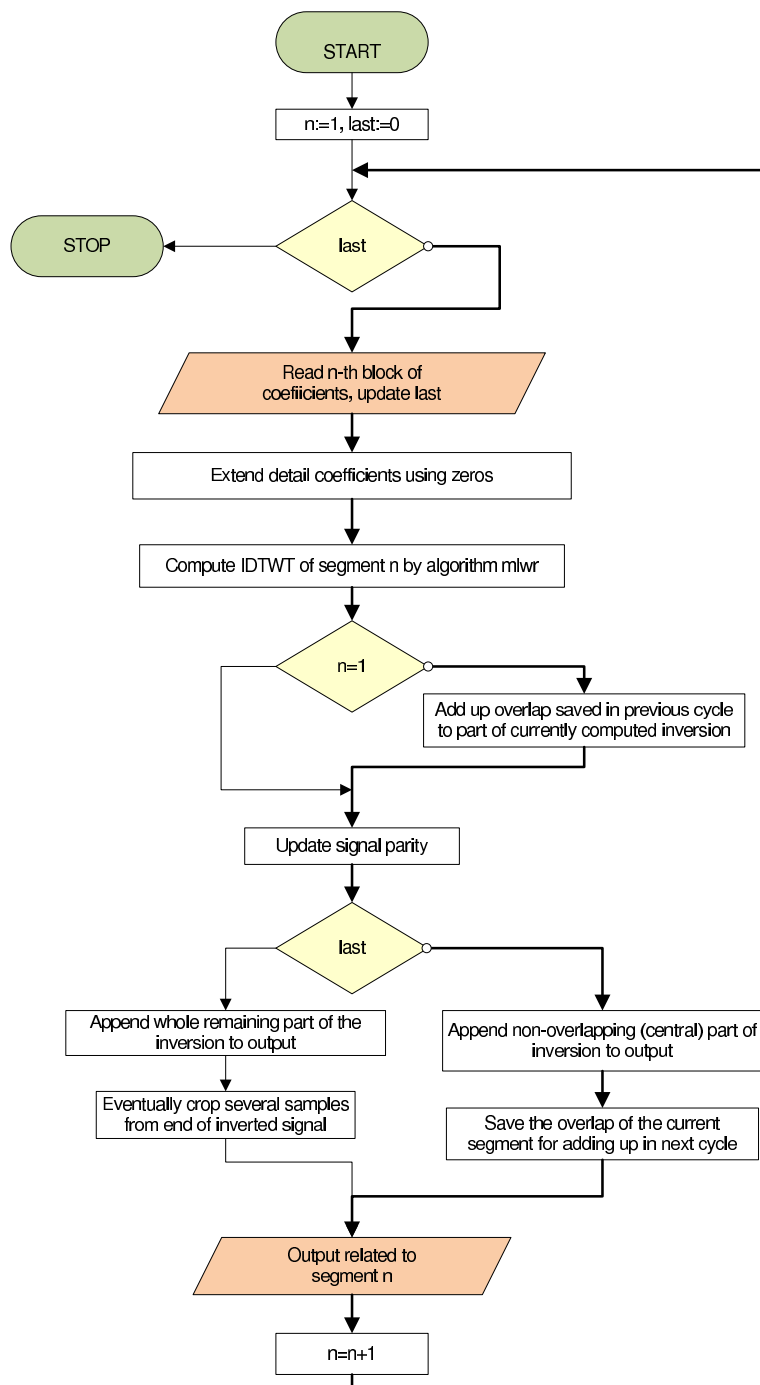


Figure 6: Flowchart of the inverse SegWT. The main loop is emphasized by the thicker line.

### 3.6 Joining Forward and Inverse Parts to Form Algorithm Capable of Real-Time Performance

The algorithms in Sec. 3.2 and 3.5 were presented separately for clarity. However, their easy integration into a simple joint loop forms a universal algorithm for any wavelet-type processing task in real time. It can be shown that in the case of  $s$  being divisible by  $2^J$  the total delay is no bigger than  $s$  samples, in other cases no bigger than  $2s$ .

**Acknowledgments** The paper was prepared within the framework of No. 102/06/P407 and No. 102/07/1303 projects of the Grant Agency of the Czech Republic and No. 1ET301710509 project of the Czech Academy of Sciences.

### References

- [1] Ch. Chrisafis and A. Ortega, Line-Based, Reduced Memory, Wavelet Image Compression. *IEEE Transactions on Image Processing*, vol. 9, no. 3, pp. 378–389, 2000.
- [2] D. Darlington, L. Daudet and M. Sandler, Digital Audio Effects in the Wavelet Domain. In *Proc. of the 5th Int. Conf. on Digital Audio Effects (DAFX-02)*, Hamburg, 2002.
- [3] W. Jiang and A. Ortega, Lifting factorization-based discrete wavelet transform architecture design. *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 11, no. 5, pp. 651–657, 2001.
- [4] Hd.O. Mota and F.H. Vasconcelos and R.M. Silva, Real-time wavelet transform algorithms for the processing of continuous streams of data, *Intelligent Signal Processing, 2005 IEEE International Workshop on*, 1–3 Sept. 2005, pp. 346–351. ISBN 0-7803-9030-x
- [5] J.H. Nealand and A.B. Bradley and M. Lech, Overlap-Save Convolution Applied to Wavelet Analysis. *IEEE Signal Processing Letters*, vol. 10, no. 2, pp. 47–49, 2003.
- [6] P. Rajmic, *Exploitation of the wavelet transform and mathematical statistics for separation signals and noise, (in Czech)*, PhD Thesis, Brno University of Technology, Brno, 2004.
- [7] G. Strang. and T. Nguyen, *Wavelets and Filter Banks*. Wellesley Cambridge Press, 1996.

# APPROXIMATING COMMON FIXED POINTS BY AN ITERATIVE PROCESS INVOLVING TWO STEPS AND THREE MAPPINGS

SAFEER HUSSAIN KHAN

**ABSTRACT.** Agarwal et al [1] introduced an iteration process which involves two steps and one mapping. They proved some results using nearly uniformly  $k$ -contractions. On the other hand, Berinde [2] introduced a new class of quasi-contractive type operators on a normed space. First we compare these two types of mappings. We then modify both Agarwal's process and mappings of Berinde to the case of three mappings keeping the number of steps same. We use this modified process first to prove some strong convergence theorems to approximate common fixed points of three quasi-contractive operators in normed spaces and then three nearly uniformly  $k$ -contractions in uniformly convex Banach spaces. This will generalize corresponding results of Berinde [2], Agarwal et al [1] and unify a number of results.

## 1. INTRODUCTION AND PRELIMINARIES

Let  $C$  be a nonempty convex subset of a normed space  $E$  and  $T : C \rightarrow C$  be a mapping. Let  $\{a_n\}$  be appropriately chosen sequence in  $(0, 1)$ . Throughout this paper,  $\mathbb{N}$  will denote the set of all positive integers,  $I$  the identity mapping on  $C$ ,  $F(T)$  the set of all fixed points of  $T$  and  $F = \bigcap_{i=1}^3 F(T_i)$ , the set of common fixed points of the mappings  $T_i : C \rightarrow C, i = 1, 2, 3$ .

The Mann iterative process [5] is defined by the sequence  $\{x_n\}$ :

$$(1.1) \quad \begin{cases} x_1 = x \in C, \\ x_{n+1} = (1 - \alpha_n) x_n + \alpha_n T x_n, \quad n \in \mathbb{N} \end{cases}$$

where  $\{\alpha_n\}$  is in  $(0, 1)$ .

---

2000 *Mathematics Subject Classification.* 47H10, 54H25 .

*Key words and phrases.* Quasi-contractive type operator, Nearly uniformly  $k$ -contraction, Iterative process, Common fixed point, Strong convergence.

The sequence  $\{x_n\}$  defined by

$$(1.2) \quad \begin{cases} x_1 = x \in C, \\ x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T y_n, \\ y_n = (1 - \beta_n)x_n + \beta_n T x_n, \quad n \in \mathbb{N} \end{cases}$$

where  $\{\alpha_n\}$  and  $\{\beta_n\}$  are in  $(0, 1)$ , is known as the Ishikawa iterative process [4].

Recently, Agarwal et al [1] introduced the following iterative process:

$$(1.3) \quad \begin{cases} x_1 = x \in C, \\ x_{n+1} = (1 - \alpha_n)T x_n + \alpha_n T y_n, \\ y_n = (1 - \beta_n)x_n + \beta_n T x_n, \quad n \in \mathbb{N} \end{cases}$$

where  $\{\alpha_n\}$  and  $\{\beta_n\}$  in  $(0, 1)$ .

Note that neither (1.1) nor (1.2) can be deduced from (1.3). They defined nearly uniformly  $k$ -contraction as a mapping  $T : C \rightarrow C$  satisfying

$$(1.4) \quad \|T^n x - T^n y\| \leq k(\|x - y\| + a_n)$$

where  $0 < k < 1$  and  $\{a_n\}$  is a sequence in  $[0, \infty)$  with  $a_n \rightarrow 0$ . Clearly every contraction is nearly uniformly  $k$ -contraction. They proved the following strong convergence result.

**Theorem 1.** *Let  $E$  be a uniformly convex Banach space and let  $C$  be its closed and convex subset. Let  $T : C \rightarrow C$  be a nearly uniformly  $k$ -contraction with a sequence  $\{a_n\}$  and  $F(T) \neq \emptyset$  such that  $\sum_{n=1}^{\infty} a_n < \infty$ . Define a sequence  $\{x_n\}$  in  $C$  as:*

$$\begin{cases} x_1 = x \in C, \\ x_{n+1} = (1 - \alpha_n)T^n x_n + \alpha_n T^n y_n, \\ y_n = (1 - \beta_n)x_n + \beta_n T^n x_n, \quad n \in \mathbb{N} \end{cases}$$

where  $\{\alpha_n\}, \{\beta_n\}$  are in  $(0, 1)$ . Then  $\{x_n\}$  converges strongly to a fixed point of  $T$ .

On the other hand, Berinde [2] introduced a new class of quasi-contractive type operators on a normed space  $E$  satisfying

$$(1.5) \quad \|Tx - Ty\| \leq \delta \|x - y\| + L \|Tx - x\|$$

for any  $x, y \in E$ ,  $0 < \delta < 1$  and  $L \geq 0$ . This class of mappings is larger than not only contractions but also Kannan mappings and Zamfirescu operators. For details, see [2].

The following comparison of the definitions (1.4) and (1.5) shows that Theorem 1 does not cover the above type of operators.

## TWO STEPS THREE MAPPINGS ITERATIVE PROCESS

**Proposition 1.** (1.4) does not imply (1.5) in general. However, if  $T$  is identity mapping or  $x$  is a fixed point of  $T$  or  $L = 0$ , then we must choose  $a_n$  identically zero.

*Proof.* From (1.4),

$$\begin{aligned}
 \|T^n x - T^n y\| &= \|T(T^{n-1}x) - T(T^{n-1}y)\| \\
 &\leq \delta \|T^{n-1}x - T^{n-1}y\| + L \|T^n x - T^{n-1}x\| \\
 &\leq \delta (\delta \|T^{n-2}x - T^{n-2}y\| + L \|T^{n-1}x - T^{n-2}x\|) \\
 &\quad + L \|T^n x - T^{n-1}x\| \\
 &= \delta^2 \|T^{n-2}x - T^{n-2}y\| + \delta L \|T^{n-1}x - T^{n-2}x\| \\
 &\quad + L \|T^n x - T^{n-1}x\| \\
 &\leq \delta^3 \|T^{n-3}x - T^{n-3}y\| + \delta^2 L \|T^{n-2}x - T^{n-3}x\| \\
 &\quad + \delta L \|T^{n-1}x - T^{n-2}x\| + L \|T^n x - T^{n-1}x\| \\
 &\quad \vdots \\
 &\leq \delta^n \|x - y\| + L \left( \begin{aligned} &\delta^{n-1} \|Tx - x\| + \delta^{n-2} \|T^2x - Tx\| + \dots \\ &+ \delta \|T^{n-1}x - T^{n-2}x\| + \|T^n x - T^{n-1}x\| \end{aligned} \right)
 \end{aligned}$$

That is,

$$(1.6) \quad \|T^n x - T^n y\| \leq \delta^n \|x - y\| + L \left( \begin{aligned} &\delta^{n-1} \|Tx - x\| + \delta^{n-2} \|T^2x - Tx\| + \dots \\ &+ \delta \|T^{n-1}x - T^{n-2}x\| + \|T^n x - T^{n-1}x\| \end{aligned} \right)$$

But

$$\begin{aligned}
 \|T^2x - Tx\| &= \|T(Tx) - Tx\| \\
 &\leq \delta \|Tx - x\| + L \|x - Tx\| \\
 &= (\delta + L) \|x - Tx\|,
 \end{aligned}$$

and

$$\begin{aligned}
 \|T^3x - T^2x\| &= \|T(T^2x) - T(Tx)\| \\
 &\leq \delta \|T^2x - Tx\| + L \|T^2x - Tx\| \\
 &= (\delta + L) \|T^2x - Tx\| \\
 &\leq (\delta + L)^2 \|x - Tx\|
 \end{aligned}$$

and so on, we get

$$\|T^n x - T^{n-1}x\| \leq (\delta + L)^{n-1} \|x - Tx\|$$

so that

$$\begin{aligned}
& \delta^{n-1} \|Tx - x\| + \delta^{n-2} \|T^2x - Tx\| + \cdots \\
& + \delta \|T^{n-1}x - T^{n-2}x\| + \|T^nx - T^{n-1}x\| \\
\leq & \left( \begin{array}{c} \delta^{n-1} + \delta^{n-2}(\delta + L) + \delta^{n-3}(\delta + L)^2 + \cdots \\ + \delta(\delta + L)^{n-2} + (\delta + L)^{n-1} \end{array} \right) \|x - Tx\| \\
= & \delta^{n-1} \left( \begin{array}{c} 1 + \left(\frac{\delta+L}{\delta}\right) + \left(\frac{\delta+L}{\delta}\right)^2 + \cdots \\ + \left(\frac{\delta+L}{\delta}\right)^{n-2} + \left(\frac{\delta+L}{\delta}\right)^{n-1} \end{array} \right) \|x - Tx\| \\
= & \delta^{n-1} \left( \begin{array}{c} 1 + \left(1 + \frac{L}{\delta}\right) + \left(1 + \frac{L}{\delta}\right)^2 + \cdots \\ + \left(1 + \frac{L}{\delta}\right)^{n-2} + \left(1 + \frac{L}{\delta}\right)^{n-1} \end{array} \right) \|x - Tx\| \\
= & \delta^{n-1} \left( \frac{1 - \left(1 + \frac{L}{\delta}\right)^n}{1 - \left(1 + \frac{L}{\delta}\right)} \right) \|x - Tx\| \\
= & \frac{\delta^n}{L} \left[ \left(1 + \frac{L}{\delta}\right)^n - 1 \right] \|x - Tx\|.
\end{aligned}$$

Hence (1.6) becomes

$$\begin{aligned}
\|T^nx - T^ny\| & \leq \delta^n \|x - y\| + L \frac{\delta^n}{L} \left[ \left(1 + \frac{L}{\delta}\right)^n - 1 \right] \|x - Tx\| \\
& = \delta^n \|x - y\| + \delta^n \left[ \left(1 + \frac{L}{\delta}\right)^n - 1 \right] \|x - Tx\| \\
& = \delta^n \left( \|x - y\| + \left[ \left(1 + \frac{L}{\delta}\right)^n - 1 \right] \|x - Tx\| \right) \\
& \leq \delta \left( \|x - y\| + \left[ \left(1 + \frac{L}{\delta}\right)^n - 1 \right] \|x - Tx\| \right).
\end{aligned}$$

Choose  $k = \delta$ . Define

$$a_n = \left[ \left(1 + \frac{L}{\delta}\right)^n - 1 \right] \|x - Tx\|.$$

If  $T$  is identity mapping or  $x$  is a fixed point of  $T$  or  $L = 0$ , then  $a_n$  is identically zero. If  $L \neq 0$ , then  $1 + \frac{L}{\delta} \in (1, \infty)$  and so  $\left(1 + \frac{L}{\delta}\right)^n$  diverges. Hence  $a_n \not\rightarrow 0$  as  $n \rightarrow \infty$ . Consequently, any mapping satisfying (1.4) does not satisfy (1.5) in general.  $\square$

Berinde [2] used the Ishikawa iterative process (1.2) to approximate fixed points of the class of operators (1.5) in a normed space. Actually, his main theorem was the following:

## TWO STEPS THREE MAPPINGS ITERATIVE PROCESS

**Theorem 2.** *Let  $C$  be a nonempty closed convex subset of a normed space  $E$ . Let  $T : C \rightarrow C$  be an operator satisfying (1.5). Let  $\{x_n\}$  be defined by the iterative process (1.2). If  $F(T) \neq \emptyset$  and  $\sum_{n=1}^{\infty} \alpha_n = \infty$ , then  $\{x_n\}$  converges strongly to a fixed point of  $T$ .*

Now note that although the iterative process used in Theorem 1 is better than the one used in Theorem 2 (see [1]) but it does not cover the type of operators used in Theorem 2. We, thus, need a modification for both of these. Keeping in mind that approximating common fixed points has a direct link with the minimization problem, see for example [6], we modify both (1.3) and (1.5) to the case of three mappings  $T_1, T_2$  and  $T_3$  as follows.

$$(1.7) \quad \begin{cases} x_1 = x \in C, \\ x_{n+1} = (1 - \alpha_n)T_1x_n + \alpha_nT_2y_n, \\ y_n = (1 - \beta_n)x_n + \beta_nT_3x_n, \quad n \in \mathbb{N} \end{cases}$$

where  $\{\alpha_n\}$  and  $\{\beta_n\}$  are in  $(0, 1)$  and

$$(1.8) \quad \max_{i \in \{1,2,3\}} \|T_i x - T_i y\| \leq \delta \|x - y\| + L \max_{i \in \{1,2,3\}} \|T_i x - x\|$$

for any  $x, y \in E, 0 < \delta < 1$  and  $L \geq 0$ .

We note that (1.7) reduces to

- (1.3) when  $T_i = T$  for all  $i = 1, 2, 3$ ,
- (1.2) when  $T_1 = I, T_2 = T_3 = T$ ,
- (1.1) when  $T_1 = T_3 = I, T_2 = T$ .

We also note that (1.8) reduces to (1.5) in any of the following cases:

- when  $T_i = T$  for all  $i = 1, 2, 3$ ,
- when any two of  $T_i, i = 1, 2, 3$  equal  $T$  and the third one is  $I$ ,
- when two of  $T_i, i = 1, 2, 3$  are identity but the third one is  $T$ .

In the rest of the paper, we use the iterative process (1.7) where the mappings satisfy (1.8) to prove a common-fixed-point-result in normed spaces. A corollary to this result will cover the case of operators defined by (1.5) using (1.3). We also obtain generalizations of some results of [2]. Moreover, we use a variant of (1.7) to prove a result for nearly uniformly  $k$ -contractions in Banach spaces thereby generalizing a result of Agarwal et al [1].

## 2. COMMON FIXED POINTS BY A TWO STEPS THREE MAPPINGS PROCESS

**2.1. Results in normed spaces.** Our first theorem deals with the iterative process (1.7) for the mappings defined in (1.8).

**Theorem 3.** *Let  $C$  be a nonempty closed convex subset of a normed space  $E$ . Let  $T_i : C \rightarrow C$ ,  $i = 1, 2, 3$  be three operators satisfying (1.8) and  $F \neq \emptyset$ . Let  $\{x_n\}$  be defined by the iterative process (1.7). If  $\{\alpha_n\}$  and  $\{\beta_n\}$  are sequences in  $(0, 1)$  such that  $\sum_{n=1}^{\infty} \alpha_n \beta_n = \infty$ , then  $\{x_n\}$  converges strongly to a point of  $F$ .*

*Proof.* Let  $w \in F$ . Then

$$\begin{aligned} \|x_{n+1} - w\| &= \|(1 - \alpha_n)T_1x_n + \alpha_nT_2y_n - w\| \\ (2.1) \quad &\leq (1 - \alpha_n)\|T_1x_n - w\| + \alpha_n\|T_2y_n - w\|. \end{aligned}$$

Since  $\|T_2y_n - w\| \leq \max_{i \in \{1, 2, 3\}} \|T_iy_n - w\|$ , therefore for  $x = w$  and  $y = y_n$ , (1.8) gives

$$(2.2) \quad \|T_2y_n - w\| \leq \delta \|y_n - w\|$$

Similarly, the choice  $x = w$  and  $y = x_n$  provides

$$(2.3) \quad \|T_3x_n - w\| \leq \delta \|x_n - w\|.$$

But

$$\begin{aligned} \|y_n - w\| &\leq (1 - \beta_n)\|x_n - w\| + \beta_n\|T_3x_n - w\| \\ &\leq (1 - \beta_n)\|x_n - w\| + \beta_n\delta\|x_n - w\| \\ (2.4) \quad &\leq (1 - \beta_n(1 - \delta))\|x_n - w\|. \end{aligned}$$

Then using of (2.1) through (2.4), we obtain

$$\begin{aligned} \|x_{n+1} - w\| &\leq (1 - \alpha_n)\|T_1x_n - w\| + \alpha_n\|T_2y_n - w\| \\ &\leq (1 - \alpha_n)\delta\|x_n - w\| + \alpha_n\delta(1 - \beta_n(1 - \delta))\|x_n - w\| \\ &= [(1 - \alpha_n)\delta + \alpha_n\delta(1 - \beta_n(1 - \delta))]\|x_n - w\| \\ &= \delta[(1 - \alpha_n + \alpha_n - \alpha_n\beta_n(1 - \delta))]\|x_n - w\| \\ &= \delta[(1 - \alpha_n\beta_n(1 - \delta))]\|x_n - w\| \end{aligned}$$

By induction,

$$\begin{aligned} \|x_{n+1} - w\| &\leq \prod_{k=1}^n [1 - (1 - \delta)\alpha_k\beta_k]\|x_1 - w\| \\ &= \|x_1 - w\| \exp\left(\sum_{k=1}^n -(1 - \delta)\alpha_k\beta_k\right) \\ &= \|x_1 - w\| \exp\left(-(1 - \delta)\sum_{k=1}^n \alpha_k\beta_k\right) \end{aligned}$$

for all  $n \in \mathbb{N}$ .



## TWO STEPS THREE MAPPINGS ITERATIVE PROCESS

Since  $0 < \delta < 1$ ,  $\alpha_n, \beta_n \in (0, 1)$  and  $\sum_{n=1}^{\infty} \alpha_n \beta_n = \infty$ , we get that

$$\limsup_{n \rightarrow \infty} \|x_n - w\| \leq \limsup_{n \rightarrow \infty} \|x_1 - w\| \exp \left( - (1 - \delta) \sum_{k=1}^n \alpha_k \beta_k \right) \leq 0.$$

Hence  $\lim_{n \rightarrow \infty} \|x_n - w\| = 0$ . Consequently  $x_n \rightarrow w \in F$ . This completes the proof.  $\square$

Although the following theorem is a corollary to our above theorem yet it is new in itself. This theorem also complements and improves a result of [1] to the case of the operators defined by (1.5) in normed spaces. Actually it is a blend of Theorems 1 and 2 and covers the case of operators defined by (1.5) using (1.3).

**Theorem 4.** *Let  $C$  be a nonempty closed convex subset of a normed space  $E$ . Let  $T : C \rightarrow C$  be an operator satisfying (1.5) and  $F(T) \neq \emptyset$ . Let  $\{x_n\}$  be defined by the iterative process (1.3). If  $\{\alpha_n\}$  and  $\{\beta_n\}$  are sequences in  $(0, 1)$  such that  $\sum_{n=1}^{\infty} \alpha_n \beta_n = \infty$ , then  $\{x_n\}$  converges strongly to a fixed point of  $T$ .*

*Proof.* Choose  $T_1 = T_2 = T_3 = T$  in Theorem 3.  $\square$

Note that the following cannot be obtained as a corollary by using the iterative process (1.3). However, we can get it by using (1.7). Note also that it is Theorem 1 of [2]. Furthermore, the following corollary generalizes Theorem 2 of [3] and the results generalized therein. Thus our result also unifies a number of results in the literature.

**Corollary 1.** *Let  $C$  be a nonempty closed convex subset of a normed space  $E$ . Let  $T : C \rightarrow C$  be an operator satisfying (1.5) and  $F(T) \neq \emptyset$ . Let  $\{x_n\}$  be defined by the iterative process (1.2). If  $\{\alpha_n\}$  and  $\{\beta_n\}$  are sequences in  $(0, 1)$  such that  $\sum_{n=1}^{\infty} \alpha_n \beta_n = \infty$ , then  $\{x_n\}$  converges strongly to a fixed point of  $T$ .*

*Proof.* Choose  $T_1 = T_3 = I$ ,  $T_2 = T$  in Theorem 3.  $\square$

Similarly, we have:

**Corollary 2.** *Let  $C$  be a nonempty closed convex subset of a normed space  $E$ . Let  $T : C \rightarrow C$  be an operator satisfying (1.5) and  $F(T) \neq \emptyset$ . Let  $\{x_n\}$  be defined by the iterative process (1.1). If  $\{\alpha_n\}$  is a sequences in  $(0, 1)$  such that  $\sum_{n=1}^{\infty} \alpha_n = \infty$ , then  $\{x_n\}$  converges strongly to a fixed point of  $T$ .*

*Proof.* Choose  $T_1 = I$ ,  $T_2 = T_3 = T$  in Theorem 3.  $\square$

**2.2. Results in uniformly convex Banach spaces.** Here we generalize Theorem 3.7 of [1] to the case of three nearly uniformly  $k$ -contractions with a sequence  $\{a_n\}$ . We define  $\{x_n\}$  in  $C$  as:

$$(2.5) \quad \begin{cases} x_1 = x \in C, \\ x_{n+1} = (1 - \alpha_n)T_1^n x_n + \alpha_n T_2^n y_n, \\ y_n = (1 - \beta_n)x_n + \beta_n T_3^n x_n, \quad n \in \mathbb{N} \end{cases}$$

where  $\{\alpha_n\}$  and  $\{\beta_n\}$  are in  $(0, 1)$ .

**Theorem 5.** *Let  $C$  be a nonempty closed convex subset of a normed space  $E$ . Let  $T_i : C \rightarrow C$ ,  $i = 1, 2, 3$  be three nearly uniformly  $k$ -contractions with a sequence  $\{a_n\}$  and  $F \neq \emptyset$  such that  $\sum_{n=1}^{\infty} a_n < \infty$ . Let  $\{x_n\}$  be defined by the iterative process (2.5). If  $\{\alpha_n\}$  and  $\{\beta_n\}$  are sequences in  $(0, 1)$ , then  $\{x_n\}$  converges strongly to a common fixed point of  $T_i$ ,  $i = 1, 2, 3$ .*

*Proof.* Let  $w \in F$ . Then

$$\begin{aligned} \|x_{n+1} - w\| &= \|(1 - \alpha_n)T_1^n x_n + \alpha_n T_2^n y_n - w\| \\ &\leq (1 - \alpha_n) \|T_1^n x_n - w\| + \alpha_n \|T_2^n y_n - w\| \\ &\leq (1 - \alpha_n)k(\|x_n - w\| + a_n) + \alpha_n k(\|y_n - w\| + a_n) \\ &= k[(1 - \alpha_n)\|x_n - w\| + \alpha_n\|y_n - w\| + a_n] \\ &\leq k \left[ (1 - \alpha_n)\|x_n - w\| + \alpha_n(1 - \beta_n)\|x_n - w\| \right. \\ &\quad \left. + \alpha_n\beta_n\|T_3^n x_n - w\| + a_n \right] \\ &\leq k \left[ (1 - \alpha_n)\|x_n - w\| + \alpha_n(1 - \beta_n)\|x_n - w\| \right. \\ &\quad \left. + k\alpha_n\beta_n\|x_n - w\| + k\alpha_n\beta_na_n + a_n \right] \\ &= k \left[ (1 - \alpha_n + \alpha_n(1 - \beta_n) + k\alpha_n\beta_n)\|x_n - w\| \right. \\ &\quad \left. + k\alpha_n\beta_na_n + a_n \right] \\ &\leq k[(1 - (1 - k)\alpha_n\beta_n)\|x_n - w\| + (k + 1)a_n] \\ &\leq k\|x_n - w\| + k(k + 1)a_n \\ &\leq \|x_n - w\| + k(k + 1)a_n \end{aligned}$$

It is well-known that if  $\{r_n\}$  and  $\{s_n\}$  are sequences of nonnegative real numbers such that  $r_{n+1} \leq r_n + s_n$  and  $\sum_{n=1}^{\infty} s_n < \infty$ , then  $\lim r_n$  exists. Thus  $\lim \|x_n - w\|$  exists. Call it  $c$ . If  $c > 0$ , then  $a_n \rightarrow 0$  together with  $\|x_{n+1} - w\| \leq k\|x_n - w\| + k(k + 1)a_n$  gives  $c \leq kc$ , a contradiction. Hence  $\lim \|x_n - w\| = 0$  and  $\{x_n\}$  converges strongly to a common fixed point of  $T_i$ ,  $i = 1, 2, 3$  as required.  $\square$

## TWO STEPS THREE MAPPINGS ITERATIVE PROCESS

**Remark.** In the above theorem, if  $T_1$  is a nearly uniformly  $k_1$ -contraction with a sequence  $\{a_n^1\}$ ,  $T_2$  is a nearly uniformly  $k_2$ -contraction with a sequence  $\{a_n^2\}$  and  $T_3$  is a nearly uniformly  $k_3$ -contraction with a sequence  $\{a_n^3\}$ , then we can choose  $a_n = \min(a_n^1, a_n^2, a_n^3)$  and  $k = \min(k_1, k_2, k_3)$  so that our result still remains valid.

Following is the Theorem 3.7 of [1] which we can obtain now by choosing  $T_1 = T_2 = T_3 = T$  in Theorem 5.

**Corollary 3.** Let  $E$  be a uniformly convex Banach space and let  $C$  be its closed and convex subset. Let  $T : C \rightarrow C$  be a nearly uniformly  $k$ -contraction with a sequence  $\{a_n\}$  and  $F(T) \neq \emptyset$  such that  $\sum_{n=1}^{\infty} a_n < \infty$ . Define a sequence  $\{x_n\}$  in  $C$  as in (1.3) where  $\{\alpha_n\}, \{\beta_n\}$  are in  $(0, 1)$ . Then  $\{x_n\}$  converges strongly to a fixed point of  $T$ .

**Acknowledgement.** The author gratefully acknowledges the support from Qatar University, Qatar to carry out this work.

## REFERENCES

- [1] R.P. Agarwal, D. O'Regan and D.R. Sahu, *Iterative construction of fixed points of nearly asymptotically nonexpansive mappings*, J. Nonlinear Convex Anal., **8**(1)2007, 61-79.
- [2] V. Berinde, *A convergence theorem for some mean value fixed point iterations procedures*, Dem.Math., **38**(1)2005, 177-184.
- [3] ———, *On the convergence of Ishikawa iteration in the class of quasi contractive operators*, Acta.Math.Univ.Comenianae, **LXXIII** (1) 2004, 119-126.
- [4] S. Ishikawa, *Fixed points by a new iteration method*, Proc.Amer.Math.Soc., **44** (1974), 147-150.
- [5] W.R. Mann, *Mean value methods in iterations*, Proc.Amer.Math.Soc., **4** (1953), 506-510.
- [6] W. Takahashi, *Iterative methods for approximation of fixed points and their applications*, J.Oper.Res.Soc. Jpn., **43**(1) (2000), 87 -108.

SAFEER HUSSAIN KHAN, DEPARTMENT OF MATHEMATICS AND PHYSICS, QATAR UNIVERSITY, DOHA 2713, STATE OF QATAR.

E-mail address: safeerhussain5@yahoo.com; safeer@qu.edu.qa

# LAWTON'S CONDITIONS ON REGULAR LOW PASS FILTERS

A. San Antolín

Department of Mathematics and Statistics, Auburn University,  
Auburn, AL., USA, 36849.

E-mail: azs0033@auburn.edu

## Abstract

For the study of  $\mathbb{Z}^n$ -periodic bounded measurable functions  $H$  which are low pass filters in an multiresolution analysis defined on  $L^2(\mathbb{R}^n)$  with a dilation given by a fixed linear invertible map  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $A(\mathbb{Z}^n) \subset \mathbb{Z}^n$  and all (complex) eigenvalues of  $A$  have modulus greater than 1, one should assume that the infinite product  $\prod_{j=1}^{\infty} |H((A^*)^{-j}\mathbf{t})|$  converges almost everywhere on  $\mathbb{R}^n$  and is  $A^*$ -locally nonzero at the origin, where  $A^*$  is the adjoint map of  $A$ . In this paper we find a condition on the regularity of  $H$  at the origin which assures that the above requirements on the infinite product hold. Moreover, depending of the regularity we assume on  $H$  we get different necessary and sufficient conditions on  $H$  to be a low pass filter in an  $A$ -MRA following the strategy of Lawton.

Keywords: Fourier transform, Hölder continuous function, Lawton's conditions, locally nonzero function, low pass filter in a multiresolution analysis.

## 1 Introduction and Definitions.

A multiresolution analysis (MRA) is a general method introduced by Mallat [21] and Meyer [22] for constructing wavelets. Afterwards, the concept of MRA was considered on  $L^2(\mathbb{R}^n)$ ,  $n \geq 1$ , (see [20],[11],[26],[27]) in a more general context, where instead of the dyadic dilation one considers the dilation given by a fixed linear invertible map  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  such that  $A(\mathbb{Z}^n) \subset \mathbb{Z}^n$  and all (complex) eigenvalues of  $A$  have modulus greater than 1. Here and further we use the same notation for the linear invertible map  $A$  and its matrix with respect to the canonical base. Given such a linear invertible map  $A$  one defines an  $A$ -MRA as a sequence of closed subspaces  $V_j$ ,  $j \in \mathbb{Z}$ , of the Hilbert space  $L^2(\mathbb{R}^n)$  that satisfies the following conditions:

- (i)  $\forall j \in \mathbb{Z}, \quad V_j \subset V_{j+1};$
- (ii)  $\forall j \in \mathbb{Z}, \quad f(\mathbf{x}) \in V_j \Leftrightarrow f(A\mathbf{x}) \in V_{j+1};$

A. San Antolín

- (iii)  $\overline{\cup_{j \in \mathbb{Z}} V_j} = L^2(\mathbb{R}^n)$ ;
- (iv) There exists a function  $\phi \in V_0$ , that is called *scaling function*, such that  $\{\phi(\mathbf{x} - \mathbf{k}) : \mathbf{k} \in \mathbb{Z}^n\}$  is an orthonormal basis for  $V_0$ .

Properties of scaling functions have been studied by several authors (see [21],[15],[9],[4],[11],[1],[8],[14],[19],[5]).

In this paper, we adopt the convention that the Fourier transform of a function  $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$  is defined by

$$\hat{f}(\mathbf{y}) = \int_{\mathbb{R}^n} f(\mathbf{x}) e^{-2\pi i \mathbf{x} \cdot \mathbf{y}} d\mathbf{x}.$$

If  $\phi$  is a scaling function of an  $A$ -MRA, observe that  $d_A^{-1}\phi(A^{-1}\mathbf{x}) \in V_{-1} \subset V_0$ , where  $d_A = |\det A|$ . By the condition (iv) we express this function in terms of the orthonormal basis  $\{\phi(\mathbf{x} - \mathbf{k}) : \mathbf{k} \in \mathbb{Z}^n\}$  as

$$d_A^{-1}\phi(A^{-1}\mathbf{x}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} a_{\mathbf{k}} \phi(\mathbf{x} - \mathbf{k}),$$

where the convergence is in  $L^2(\mathbb{R}^n)$  and  $\{a_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^n} \in l^2$ . Taking the Fourier transform, we obtain

$$\hat{\phi}(A^* \mathbf{t}) = H(\mathbf{t}) \hat{\phi}(\mathbf{t}) \quad \text{a.e. on } \mathbb{R}^n$$

where  $A^*$  is the adjoint map of  $A$  and

$$H(\mathbf{t}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} a_{\mathbf{k}} e^{-2\pi i \mathbf{k} \cdot \mathbf{t}}$$

is a  $\mathbb{Z}^n$ -periodic function which is called *low pass filter* associated with the scaling function  $\phi$ , or shortly *low pass filter*. We study the problem of when a given measurable function  $H$  is a low pass filter in an  $A$ -MRA assuming some regularity on  $H$ .

Before formulating our results let us introduce some notation and definitions.

Let  $\{\mathbf{e}_i\}_{i=1}^n$  be the natural basis of  $\mathbb{R}^n$ ,  $\mathbb{T}^n = \mathbb{R}^n / \mathbb{Z}^n$  and if we set  $f \in L^2(\mathbb{T}^n)$  we will understand that  $f$  is defined on the whole space  $\mathbb{R}^n$  as a  $\mathbb{Z}^n$ -periodic function. With some abuse of the notation we consider also that  $\mathbb{T}^n$  is the unit cube  $[0, 1)^n$ .

We will denote  $B_r = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| < r\}$ . For a set  $E \subset \mathbb{R}^n$  and a point  $\mathbf{x} \in \mathbb{R}^n$  we will write  $\mathbf{x} + E = \{\mathbf{x} + \mathbf{y} : \mathbf{y} \in E\}$ . The Lebesgue measure of a measurable set  $E \subset \mathbb{R}^n$  will be denoted by  $|E|_n$  and by  $\chi_E$  the characteristic function of the set  $E$  i.e.  $\chi_E(\mathbf{t})$  takes the value 1 if  $\mathbf{t} \in E$  and 0 otherwise.

Given  $N \in \{1, 2, \dots\}$ , the set of  $N$  times differentiable functions  $f : \mathbb{R}^n \rightarrow \mathbb{C}$  will be denoted by  $C^N(\mathbb{R}^n)$ .

We will say that a measurable function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  is *Hölder continuous* at  $\mathbf{x}_0 \in \mathbb{R}^n$  (cf. [24]) if there exist an open neighborhood of  $\mathbf{x}_0$ ,  $U \subset \mathbb{R}^n$ , and constants  $C, \alpha > 0$  such that

$$|f(\mathbf{y}) - f(\mathbf{x}_0)| \leq C|\mathbf{y} - \mathbf{x}_0|^\alpha, \quad \forall \mathbf{y} \in U. \quad (1)$$

A. San Antolín

If  $\alpha=1$ ,  $f$  is said to be *Lipschitz continuous at  $\mathbf{x}_0$* .

In [5] the following definitions were introduced.

**Definition 1.** We will say that  $\mathbf{x} \in \mathbb{R}^n$  is a point of  $A$ -density for a set  $E \subset \mathbb{R}^n$ ,  $|E|_n > 0$ , if for any  $r > 0$

$$\lim_{j \rightarrow \infty} \frac{|E \cap (A^{-j}B_r + \mathbf{x})|_n}{|A^{-j}B_r|_n} = 1.$$

**Definition 2.** Let  $f : \mathbb{R}^n \rightarrow \mathbb{C}$  be a measurable function. We say that  $\mathbf{x} \in \mathbb{R}^n$  is a point of  $A$ -approximate continuity of the function  $f$  if there exists  $E \subset \mathbb{R}^n$ ,  $|E|_n > 0$ , such that  $\mathbf{x}$  is a point of  $A$ -density for the set  $E$  and

$$\lim_{\substack{\mathbf{y} \rightarrow \mathbf{x} \\ \mathbf{y} \in E}} f(\mathbf{y}) = f(\mathbf{x}).$$

**Definition 3.** A measurable function  $f : \mathbb{R}^n \rightarrow \mathbb{C}$  is said to be  *$A$ -locally nonzero* at a point  $\mathbf{x} \in \mathbb{R}^n$  if for any  $\varepsilon, r > 0$  there exists  $j \in \mathbb{N}$  such that

$$|\{ \mathbf{y} \in A^{-j}B_r + \mathbf{x} : f(\mathbf{y}) = 0 \}|_n < \varepsilon |A^{-j}B_r|_n.$$

For a given  $\phi \in L^2(\mathbb{R}^n)$ , set

$$\Phi_\phi(\mathbf{t}) = \sum_{\mathbf{k} \in \mathbb{Z}^n} |\widehat{\phi}(\mathbf{t} + \mathbf{k})|^2. \quad (2)$$

If  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$  is a linear invertible map such that  $A(\mathbb{Z}^n) \subset \mathbb{Z}^n$  and all (complex) eigenvalues of  $A$  have modulus greater than 1, the quotient group  $\mathbb{Z}^n/A(\mathbb{Z}^n)$  is well defined, then we will denote by  $\Delta_A \subset \mathbb{Z}^n$  a full collection of representatives of the cosets of  $\mathbb{Z}^n/A(\mathbb{Z}^n)$ . Recall that there are exactly  $d_A$  cosets (see [11] and [27, p. 109]).

Let us fix  $\Delta_{A^*} = \{\mathbf{p}_i\}_{i=0}^{d_{A^*}-1}$ , where  $\mathbf{p}_0 = \mathbf{0}$ .

Since  $A^*$  is a linear invertible map such that all (complex) eigenvalues of  $A^*$  have modulus greater than 1 (cf. [27, p. 122], [2]) there exist  $K > 0$  and  $0 < \beta < 1$  such that

$$|(A^*)^{-j}\mathbf{t}| \leq K\beta^j|\mathbf{t}| \quad \forall j \in \{1, 2, \dots\}. \quad (3)$$

Given  $H \in L^\infty(\mathbb{T}^n)$  the following continuous linear operator  $P : L^1(\mathbb{T}^n) \rightarrow L^1(\mathbb{T}^n)$ :

$$Pf(\mathbf{t}) = \sum_{i=0}^{d_A-1} |H((A^*)^{-1}(\mathbf{t} + \mathbf{p}_i))|^2 f((A^*)^{-1}(\mathbf{t} + \mathbf{p}_i))$$

is well defined. This operator was first introduced by M. Bownik [2] as a generalization of the analogous operator introduced by W. Lawton [18] for dyadic dilations.

A. San Antolín

## 2 History References.

A. Cohen [6] gave the first necessary and sufficient conditions for a trigonometric polynomial  $H$  to be a low pass filter of an MRA on  $L^2(\mathbb{R})$ . Those conditions were extended for differentiable functions by E. Hernández and G. Weiss [14] and for the Hölder continuous functions by R. F. Gundy [12]. Furthermore, Cohen's approach was studied by M. Papadakis, H. Sikić and G. Weiss [24] to low pass filters that are Hölder continuous at the origin.

At the same time as Cohen's condition appeared, W. Lawton [17] gave another sufficient condition of a different nature when  $H$  is a trigonometric polynomial. The necessity of Lawton's condition was settled in 1990 by both A. Cohen (see [7]) and W. Lawton [18], independently, (see [9, p. 182–193]).

For our general case when an MRA is defined on  $L^2(\mathbb{R}^n)$ ,  $n \geq 1$ , and for dilations given by a map  $A$  as above described, a generalization of Cohen's conditions for low pass filters associated with characteristic scaling functions was proved by K. Gröchening and W. R. Madych [11] and by W. R. Madych [20]. Afterwards, a generalization of Cohen's and Lawton's conditions were obtained by M. Bownik [2] where the results were presented with more general assumptions about the regularity of low pass filters. Other necessary and sufficient conditions on trigonometric polynomial low pass filters appeared in the paper by J. C. Lagarias and Y. Wang [16].

The problem of characterization of low pass filters of an MRA was posed in the book by E. Hernández and G. Weiss [14].

Characterizations of low pass filters for an MRA on  $L^2(\mathbb{R})$  and the dyadic dilations are already known, see the papers by M. Papadakis, H. Sikić and G. Weiss [24] and by V. Dobrić, R. F. Gundy and P. Hitczenko [10]. Afterwards, R. F. Gundy [13] addressed the same question when the condition (iv) in the definition of MRA is relaxed by assuming that  $\{\phi(x - k) : k \in \mathbb{Z}\}$  is a Riesz basis for  $V_0$ . The author [25] proved another necessary and sufficient condition on low pass filters following the strategy of Lawton. In fact, that condition was even presented on low pass filters  $H$  in an  $A$ -MRA defined on  $L^2(\mathbb{R}^n)$ . Such a condition is written below.

Let  $\mathbf{H}_A$  be the class of all functions  $H \in L^\infty(\mathbb{T}^n)$  such that the infinite product  $\prod_{j=1}^{\infty} |H((A^*)^{-j}\mathbf{t})|$  converges almost everywhere on  $\mathbb{R}^n$  and is  $A^*$ -locally nonzero at the origin.

Moreover, let  $\Pi_A$  be the class of all measurable functions on  $\mathbb{R}^n$  such that  $f(\mathbf{0}) = 1$ ,  $0 \leq f(\mathbf{t}) \leq 1$  a.e. on  $\mathbb{R}^n$  and the origin is a point of  $A^*$ -approximate continuity of  $f$ .

**Theorem A.** *Let  $H \in \mathbf{H}_A$ . Then the following conditions are equivalent:*

- A) *The function  $|H|$  is a low pass filter associated with a scaling function  $\theta$  of an  $A$ -MRA where  $\widehat{\theta}(\mathbf{t}) := \prod_{j=1}^{\infty} |H((A^*)^{-j}\mathbf{t})|$ .*
- B) *The only function  $f \in L^1(\mathbb{T}^n) \cap \Pi_A$  invariant under the operator  $P$  is the function  $f \equiv 1$ .*

To give a complete characterization of all low pass filters associated with scaling functions, we need also the following remark done in [25].

A. San Antolín

**Remark A.** A measurable function  $H$  is a low pass filter of an  $A$ -MRA if and only if  $|H|$  is a low pass filter of some  $A$ -MRA.

Observe that it is not always easy to check whether a function  $H \in L^\infty(\mathbb{T}^n)$  belongs to the class  $\mathbf{H}_A$ . In this paper we give a condition on the regularity at the origin of the function  $H$  which assures that  $H \in \mathbf{H}_A$ . Moreover, depending on the regularity we assume on  $H$  we prove other necessary and sufficient conditions on  $H$  to be a low pass filter in an  $A$ -MRA following the strategy of Lawton. Those conditions do not appear in the literature and are new even for low pass filters in an MRA defined on  $L^2(\mathbb{R})$  with the dyadic dilation.

### 3 Main Results

We prove the following results.

**Lemma 1.** Let  $H \in L^\infty(\mathbb{T}^n)$  be a function such that  $|H(\mathbf{0})| = 1$ ,  $|H|$  is Hölder continuous at the origin and

$$\sum_{i=0}^{d_A-1} |H(\mathbf{t} + (A^*)^{-1}\mathbf{p}_i)|^2 = 1 \quad \text{a.e. on } \mathbb{R}^n, \quad (4)$$

then  $H \in \mathbf{H}_A$ .

In order not to repeat conditions let us introduce the two following classes of measurable functions.

$$\Upsilon_A = \{f \in L^1(\mathbb{T}^n) : f(\mathbf{0}) = 1, \quad f \text{ is continuous at the origin}$$

$$\text{and } 0 \leq f(\mathbf{t}) \leq 1, \text{ a.e. on } \mathbb{R}^n\}.$$

If a function  $f \in \Upsilon_A$  is also differentiable at the origin, we will say that  $f$  belongs to the class  $\Lambda_A$ .

**Theorem 1.** Let  $H$  be a measurable function such that  $|H(\mathbf{0})| = 1$ ,  $|H|$  is  $\mathbb{Z}^n$ -periodic continuous and Hölder continuous at the origin. Then the following conditions are equivalent:

- I) The function  $H$  is a low pass filter in an  $A$ -MRA.
- II) The only function  $f \in \Upsilon_A$  invariant under the operator  $P$  is the function  $f \equiv 1$ .

**Theorem 2.** Let  $H$  be a measurable function such that  $|H(\mathbf{0})| = 1$  and  $|H|$  is a  $\mathbb{Z}^n$ -periodic differentiable function. Then the following conditions are equivalent:

- 1) The function  $H$  is a low pass filter in an  $A$ -MRA.
- 2) The only function  $f \in \Lambda_A$  invariant under the operator  $P$  is the function  $f \equiv 1$ .



A. San Antolín

## 4 Proof of Lemma 1

*Proof of Lemma 1.* First of all, we prove that the infinite product

$$\prod_{j=1}^{\infty} |H((A^*)^{-j}\mathbf{t})| \quad (5)$$

converges almost everywhere on  $\mathbb{R}^n$  to a well defined measurable function.

According to the condition (4) there exists a measurable set  $E \subset \mathbb{R}^n$ ,  $|E|_n = 0$ , such that  $0 \leq |H(\mathbf{t})| \leq 1$  for every  $\mathbf{t} \in \mathbb{R}^n \setminus E$ . Let  $F = \cup_{k=-\infty}^{\infty} (A^*)^k E$ .

Given  $\varepsilon > 0$ , we set two positive numbers  $r, R > 0$  such that  $\cup_{k=0}^{\infty} (A^*)^{-k} B_r \subset B_R \subset U$  and

$$CK^{\alpha} R^{\alpha} \sum_{j=1}^{\infty} \beta^{j\alpha} < \varepsilon,$$

where  $U$  is the open neighborhood in the definition of Hölder continuous at the origin and  $C, \alpha, K$  and  $\beta$  are the corresponding constants in the inequalities (1) and (3). Let  $S = \cup_{k=0}^{\infty} (A^*)^{-k} B_r$ .

If  $\mathbf{t} \in S \setminus F$ , then for every  $j \in \{1, 2, \dots\}$

$$0 \leq 1 - |H((A^*)^{-j}\mathbf{t})| \leq CK^{\alpha} \beta^{j\alpha} |\mathbf{t}|^{\alpha} \leq CK^{\alpha} \beta^{j\alpha} R^{\alpha}.$$

Thus, for every  $J \in \{2, 3, \dots\}$

$$\begin{aligned} 0 &\leq 1 - \prod_{j=1}^J |H((A^*)^{-j}\mathbf{t})| \\ &\leq 1 - \prod_{j=2}^J |H((A^*)^{-j}\mathbf{t})| + \prod_{j=2}^J |H((A^*)^{-j}\mathbf{t})| 1 - |H((A^*)^{-1}\mathbf{t})| \\ &\leq CK^{\alpha} R^{\alpha} \sum_{j=1}^J \beta^{j\alpha} \leq CK^{\alpha} R^{\alpha} \sum_{j=1}^{\infty} \beta^{j\alpha} < \varepsilon. \end{aligned}$$

Letting  $J \rightarrow \infty$  we obtain

$$1 - \varepsilon \leq \prod_{j=1}^{\infty} |H((A^*)^{-j}\mathbf{t})| \leq 1, \quad \forall \mathbf{t} \in S \setminus F. \quad (6)$$

Furthermore, given  $\mathbf{t} \in \mathbb{R}^n \setminus F$  there exists  $N \in \{1, 2, \dots\}$  such that  $(A^*)^{-N}\mathbf{t} \in S \setminus F$ , then

$$\prod_{j=1}^N |H((A^*)^{-j}\mathbf{t})| \prod_{j=1}^{\infty} |H((A^*)^{-j}((A^*)^{-N}\mathbf{t}))|$$

converges. So, we conclude that the infinite product (5) converges a.e. on  $\mathbb{R}^n$  to a well defined measurable function.

Finally, by (6),  $\hat{\theta}$  is  $A^*$ -locally nonzero at the origin.  $\square$

A. San Antolín

## 5 Proof of Theorem 1 and Theorem 2.

In the proof of Theorems 1 and 2 we will need the following results.

The following characterization of scaling functions in a multiresolution analysis was given in [5].

**Theorem B.** *Let  $\phi \in L^2(\mathbb{R}^n)$ . Then the following conditions are equivalent:*

- (A) *The function  $\phi$  is a scaling function of an A-MRA;*
- (B) *( $\alpha$ ) The function  $\hat{\phi}$  is  $A^*$ -locally nonzero at the origin;*  
*( $\beta$ )  $\Phi_\phi(\mathbf{t}) = 1$  a.e. on  $\mathbb{T}^n$ ;*  
*( $\gamma$ ) There exists a  $\mathbb{Z}^n$ -periodic function,  $H \in L^\infty(\mathbb{T}^n)$ ,*  
 *$|H(\mathbf{t})| \leq 1$  a.e. on  $\mathbb{R}^n$ , such that*

$$\hat{\phi}(A^*\mathbf{t}) = H(\mathbf{t})\hat{\phi}(\mathbf{t}) \quad \text{a.e. on } \mathbb{R}^n;$$

- (C) *( $\alpha^*$ ) Setting  $|\hat{\phi}(\mathbf{0})| = 1$ , the origin is a point of  $A^*$ -approximate continuity of  $|\hat{\phi}|$ ; and the conditions ( $\beta$ ) and ( $\gamma$ ) hold.*

**Proposition A.** *Let  $H \in L^\infty(\mathbb{T}^n)$  be a function such that (4) holds. If the infinite product  $\prod_{j=1}^\infty |H((A^*)^{-j}\mathbf{t})|$  converges almost everywhere then*

- a) *the function  $\hat{\theta}(\mathbf{t}) := \prod_{j=1}^\infty |H((A^*)^{-j}\mathbf{t})|$  belongs to  $L^2(\mathbb{R}^n)$  and*  

$$\|\hat{\theta}\|_{L^2(\mathbb{R}^n)} \leq 1;$$
- b)  *$\Phi_\theta(\mathbf{t}) \leq 1$  a.e. on  $\mathbb{R}^n$ ;*
- c)  *$\Phi_\theta$  is a fix point for the operator  $P$ ,*

where the function  $\theta$  is defined by  $\hat{\theta}(\mathbf{t}) := \prod_{j=1}^\infty |H((A^*)^{-j}\mathbf{t})|$ .

In the above proposition, the condition a) was proved by M. Bownik [2] (cf. [9],[14]), the condition b) was proved in the proof of main result in [25] and the condition c) also was proved in [2].

**Remark B.** If in Proposition A we add the hypotheses:  $|H(\mathbf{0})| = 1$  and  $|H|$  is a  $\mathbb{Z}^n$ -periodic continuous function and also is Hölder continuous at the origin, then the function  $\hat{\theta}$  is continuous and  $\hat{\theta}(\mathbf{0}) = 1$ .

The following result was proved by M. Bownik [2].

**Proposition B.** *Assume that a  $\mathbb{Z}^n$ -periodic function  $H$  satisfying (4) and  $|H(\mathbf{0})| = 1$  is of class  $C^N(\mathbb{R}^n)$  for some  $N = 1, 2, \dots$ . Then the function  $\hat{\theta}(\mathbf{t}) = \prod_{j=1}^\infty |H((A^*)^{-j}\mathbf{t})|$  is also of class  $C^N(\mathbb{R}^n)$  and  $\hat{\theta}(\mathbf{0}) = 1$ .*

*Proof of Theorem 1.* First of all, we will prove the implication **I**)  $\implies$  **II**). According to Remark A, since  $H$  is a low pass filter in an A-MRA then  $|H|$  is a low pass filter in some A-MRA. Thus, because  $\Upsilon_A \subset L^1(\mathbb{R}^n) \cap \Pi_A$  we finish the proof applying the condition **B**) in Theorem A.

Let us prove the implication **II**)  $\implies$  **I**). According to Lemma 1 the infinite product

$$\hat{\theta}(\mathbf{t}) = \prod_{j=1}^\infty |H((A^*)^{-j}\mathbf{t})|$$

A. San Antolín

converges almost everywhere on  $\mathbb{R}^n$  to a well defined  $A^*$ -locally nonzero measurable function. Observe that  $\hat{\theta} \in L^2(\mathbb{R}^n)$  by the condition a) in Proposition A, and in addition,  $\hat{\theta}(A^*\mathbf{t}) = |H(\mathbf{t})|\hat{\theta}(\mathbf{t})$  a.e. on  $\mathbb{R}^n$ .

Let  $\theta$  be the function defined by  $\hat{\theta}$  and consider the function  $\Phi_\theta$  given by (2), then the condition c) in Proposition A tells us that  $\Phi_\theta$  is a fix point for the operator  $P$ .

If we prove that the function  $\Phi_\theta$  belongs to  $\Upsilon_A$ , then by the condition **II**) in Theorem 1 we will have that  $\Phi_\theta(\mathbf{t}) = 1$  a.e. on  $\mathbb{T}^n$ . Hence according to Theorem B the function  $\theta$  is a scaling function of an  $A$ -MRA with associated low pass filter  $|H|$ .

Obviously,  $\Phi_\theta$  is a  $\mathbb{Z}^n$ -periodic function and  $0 \leq \Phi_\theta(\mathbf{t})$ . We do not write “a.e.” in the above inequality because according to Remark B, the function  $\hat{\theta}$  is a continuous. So, by the same reason and due to the condition c) in Proposition A,  $\Phi_\theta(\mathbf{t}) \leq 1$  holds. Moreover, since  $\hat{\theta}$  is a continuous function and  $\hat{\theta}(\mathbf{0}) = 1$ , the inequalities  $\hat{\theta}(\mathbf{t}) \leq \Phi_\theta(\mathbf{t}) \leq 1$  yield that  $\Phi_\theta(\mathbf{0}) = 1$  and the origin is a point of continuity of  $\Phi_\theta$ . Therefore,  $\Phi_\theta \in \Upsilon_A$ .

Finally, applying Remark A the proof of Theorem 1 will be finished.  $\square$

*Proof of Theorem 2.* In an analogous way that the proof of the implication **I**)  $\implies$  **II**) in Theorem 1 we can prove the implication **1**)  $\implies$  **2**) in Theorem 2.

To prove the implication **2**)  $\implies$  **1**), let  $\hat{\theta}(\mathbf{t}) = \prod_{j=1}^{\infty} |H((A^*)^{-j}\mathbf{t})|$  and repeating the schema of the proof of **II**)  $\implies$  **I**) in Theorem 1, it is enough if we prove that  $\Phi_\theta \in \Lambda_A$ . From that proof we know that  $\Phi_\theta \in \Upsilon_A$ , then it remains to prove that  $\Phi_\theta$  is differentiable at the origin.

Let us check that the partial derivatives of  $\Phi_\theta$  at the origin exist and are zero. According to Proposition B the function  $\hat{\theta}$  is differentiable and  $\hat{\theta}(\mathbf{0}) = 1$ . Thus using the inequalities  $\hat{\theta}(\mathbf{t}) \leq \Phi_\theta(\mathbf{t}) \leq 1$  for every  $\mathbf{t} \in \mathbb{R}^n$  (see the proof of Theorem 1) we obtain that  $\Phi_\theta(\mathbf{0}) = 1$  and also

$$\limsup_{h \rightarrow 0} \left| \frac{\Phi_\theta(\mathbf{0}) - \Phi_\theta(\mathbf{0} + h\mathbf{e}_i)}{h} \right| \leq \limsup_{h \rightarrow 0} \frac{(\hat{\theta}(\mathbf{0}))^2 - (\hat{\theta}(\mathbf{0} + h\mathbf{e}_i))^2}{|h|} = 0,$$

where the equality is true due to the function  $(\hat{\theta})^2$  is differentiable and it takes a maximum value at the origin. Furthermore,

$$\limsup_{\mathbf{t} \rightarrow \mathbf{0}} \frac{|\Phi_\theta(\mathbf{0}) - \Phi_\theta(\mathbf{t})|}{|\mathbf{t}|} \leq \limsup_{\mathbf{t} \rightarrow \mathbf{0}} \frac{(\hat{\theta}(\mathbf{0}))^2 - (\hat{\theta}(\mathbf{t}))^2}{|\mathbf{t}|} = 0.$$

Therefore,  $\Phi_\theta$  is a differentiable function at the origin.  $\square$

## References

- [1] C.Boor, R.DeVore, A.Ron; On the construction of multivariate (pre)wavelets, *Constr. Approx.* 9,123–166(1993).

A. San Antolín

- [2] M.Bownik; Tight frames of multidimensional wavelets, *Dedicated to the memory of Richard J. Duffin. J. Fourier Anal. Appl.* 3, no. 5, 525–542(1997).
- [3] A.Bruckner; *Differentiation of real functions*, Lecture Notes in Mathematics, 659, Springer, Berlin, 1978.
- [4] C.K.Chui; *An Introduction to Wavelets*, Academic Press, Inc. 1992.
- [5] P.Cifuentes, K.S.Kazarian, A.San Antolín; Characterization of scaling functions in a multiresolution analysis, *Proc. Amer. Math. Soc.* 133, No. 4, 1013–1023(2005).
- [6] A.Cohen; Ondelettes, analyses multirésolutions et filtres miroirs en quadrature, *Ann. Inst. H. Poincaré, Anal. non linéaire* 7, no. 5, 439–459(1990).
- [7] A.Cohen, I.Daubechies, J.C.Feauveau; Biorthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* 45, no. 5, 485–560(1992).
- [8] S.Dahlke, W.Dahmen and V.Latour; Smooth refinable functions and wavelets obtained by convolution products. *Appl. Comput. Harmon. Anal.* 2, no. 1, 68–84(1995).
- [9] I.Daubechies; *Ten lectures on wavelets*, SIAM, Philadelphia, 1992.
- [10] V.Dobrić, R.F.Gundy, P.Hitczenko; Characterizations of orthonormal scale functions: a probabilistic approach, *J. Geom. Anal.* 10, no. 3, 417–434(2000).
- [11] K.Gröchening, W.R.Madych; Multiresolution analysis, Haar bases and self-similar tilings of  $R^n$ , *IEEE Trans. Inform. Theory*, 38(2), 556–568(1992).
- [12] R.F.Gundy; Two remarks concerning wavelets: Cohen's criterion for low-pass filters and Meyer's theorem on linear independence *The functional and harmonic analysis of wavelets and frames (San Antonio, TX, 1999)*, 249–258, Contemp. Math., 247, Amer. Math. Soc., Providence, RI, 1999.
- [13] R.F.Gundy; Low-pass filters, martingales, and multiresolution analyses, *Appl. Comput. Harmon. Anal.* 9, no. 2, 204–219(2000).
- [14] E.Hernández and G.Weiss; *A first course on Wavelets*, CRC Press, Inc. 1996.
- [15] R.Q.Jia and C.A.Micchelli; Using the refinement equations for the construction of pre-wavelets. II. Powers of two. *Curves and surfaces (Chamonix-Mont-Blanc, 1990)*, 209–246, Academic Press, Boston, MA, 1991.
- [16] J.C.Lagarias, Y.Wang; Orthogonality criteria for compactly supported refinable functions and refinable function vectors, *J. Fourier Anal. Appl.* 6, no. 2, 153–170(2000).
- [17] W.M.Lawton; Tight frames of compactly supported affine wavelets, *J. Math. Phys.* 31, no. 8, 1898–1901(1990).

A. San Antolín

- [18] W.M.Lawton; Necessary and sufficient conditions for constructing orthonormal wavelet bases, *J. Math. Phys.* 32, no. 1, 57–61(1991).
- [19] R.A.Lorentz, W.R.Madych, A.Sahakian; Translation and dilation invariant subspaces of  $L^2(\mathbf{R})$  and multiresolution analyses, *Applied and Computational Harmonic Analysis* 5, no. 4, 375–388(1998).
- [20] W.R.Madych; Some elementary properties of multiresolution analyses of  $L^2(R^d)$ , *Wavelets - a tutorial in theory and applications*, Ch. Chui ed., Academic Press, 259–294(1992).
- [21] S.Mallat; Multiresolution approximations and wavelet orthonormal bases for  $L^2(R)$ , *Trans. of Amer. Math. Soc.*, 315, 69–87(1989).
- [22] Y.Meyer; *Ondelettes et opérateurs. I*, Hermann, Paris (1990) [ English Translation: *Wavelets and operators*, Cambridge University Press, (1992).]
- [23] I.P.Nathanson; *Theory of functions of a real variable*, London, vol. I, 1960.
- [24] M.Papadakis, H.Sikić, G.Weiss; The characterization of low pass filters and some basic properties of wavelets, scaling functions and related concepts, *J. Fourier Anal. Appl.* 5, no. 5, 495–521(1999).
- [25] A.San Antolín; Characterization of low pass filters in a multiresolution analysis (to appear in “*Studia Mathematica*”)
- [26] R.Strichartz; Construction of orthonormal wavelets, *Wavelets: mathematics and applications*, Stud. Adv. Math., CRC, Boca Raton, FL, 23–50(1994).
- [27] P.Wojtaszczyk; *A mathematical introduction to wavelets*, London Mathematical Society, Student Texts 37, 1997.

# Strip-saturation Model Solution for Piezoelectric Strip $\sim$ by Quadratically Varying Electric Displacement

<sup>1</sup>R. R. Bhargava and <sup>2</sup>Amit Setia

Department of Mathematics

Indian Institute of Technology Roorkee

Roorkee, 247667, India

e-mail: <sup>1</sup>rajrbfma@iitr.ernet.in, <sup>2</sup>setiadma@iitr.ernet.in

## Abstract

A crack arrest model is proposed for a cracked poled infinitely long and composite narrow piezoceramic strip. The finite crack is symmetrically situated and oriented longitudinally with respect to the edges of the strip. Uniform anti-plane shear stress or strains and in-plane normal electrical displacement applied on the finite distant edges of the strip. Consequently strip yields both mechanically and electrically. Under the assumption that the strip is electrically more brittle, an electrical singularity is encountered first. It is at this level the investigations are carried. To stop the crack from further electrical polarization the rims of the developed saturation zones are prescribed quadratically varying, normal, cohesive, in-plane saturation limit electrical displacement. The load required to arrest the developed saturation zone is assessed. In-plane electrical crack opening displacement, electrical crack growth rate's expression are obtained. Case study has been presented for BaTiO<sub>3</sub>, PZT-4 and PZT-5H strips.

**Keywords:** Piezoelectric strip, strip-saturation model, crack opening displacement, crack growth rate, saturation zone

## 1 INTRODUCTION

Gao and group [8, 7] established that local energy release rate for a piezoelectric crack with electrical yielding confined to a strip in front of the crack. It is independent of the yielding parameters and can be fully determined from a linear piezoelectric crack analysis. They further investigated the effects of electrical yielding on a finite crack lying perpendicular or parallel to the poling axis of an infinite poled piezoelectric ceramic medium. A crack perpendicular to the poling axes in a general poled ferroelectric is discussed by Ru [4] for the implications of the strip-saturation model for a electric field inducing crack. He [5] also conducted the studies for mixed boundary value problem and obtained

near crack tip field for a conducting crack parallel/perpendicular to the poling axis using based on a strip-saturation model. Wang and Zhang [1] discussed an electric strip-saturation model for fracture prediction of piezoceramics containing electrically impermeable cracks. Wang and Mai [2] investigated the fracture behavior of a cracked piezoceramic medium under transient electromechanical loads. The work on cracked piezoelectric strip was started by Shindo et al. [12]. They used the theory of linear piezoelectricity to solve the electroelastic problems of a finite crack in an orthotropic piezoelectric strip. Fourier integral transform technique was used to reduce the problem to solve a pair of dual integral equations. They [13] extended the work to study the singular stress and electric field in an orthotropic piezoelectric ceramic strip containing a Griffith crack under longitudinal shear. Li [3] analyzed the problem of a finite crack in a functionally graded material strip under an antiplane mechanical and in-plane electrical loading. In this case elastic stiffness, piezoelectric constants, and dielectric permittivity were taken to vary along the thickness of the strip. Li [9] examined the strip-saturation model for piezoelectric crack in permeable environment to analyze fracture toughness of a piezoelectric ceramics. In this study a permeable crack was modeled as a vanishing thin but finite rectangular slit with surface charge deposited along a crack surface.

## 2 METHODOLOGY

As is well-known out-of-plane displacement problem along  $xoy$ -plane may be defined as

$$u_x(x, y, z) = u_y(x, y, z) \text{ and } u_z(x, y, z) = u_z(x, y) \quad (1)$$

where  $u_i (i = x, y, z)$  define the displacement components along  $x, y$  and  $z$ -directions. Similarly an in-plane electric field problem for  $xoy$ -plane is defined as

$$E_x(x, y, z) = E_x(x, y), E_y(x, y, z) = E_y(x, y) \text{ and } E_z(x, y, z) = 0 \quad (2)$$

where  $E_i, (i = x, y, z)$  denotes the electric field component along  $z$ -direction.

Consequently linear piezoelectric theory the constitutive equations may be written as

$$\sigma_{xz} = c_{44}u_{z,x} + e_{15}\phi_{,x} \quad (3)$$

$$\sigma_{yz} = c_{44}u_{z,y} + e_{15}\phi_{,y} \quad (4)$$

$$D_x = e_{15}u_{z,x} - \epsilon_{11}\phi_{,x} \quad (5)$$

$$D_y = e_{15}u_{z,y} - \epsilon_{11}\phi_{,y} \quad (6)$$

where  $\sigma_{iz}, D_i (i = x, y)$  denote the shear stress component, electric displacement component. A comma after function denotes its partial differentiation with respect to the argument following it.  $c_{44}, e_{15}$  and  $\epsilon_{11}$  denote elastic piezoelectric and dielectric constants respectively.

The gradient equations reduce to

$$\gamma_{iz} = u_{z,i} \quad (7)$$

$$E_i = -\phi_{,i} \quad (8)$$

where  $i = x, y$ .

Stress equilibrium equation in absence of body forces are given by

$$\sigma_{ij,j} = 0 \quad (9)$$

where  $i, j = x, y, z$ . Electrical displacement equation in absence of body electric charge may be written as

$$D_{i,i} = 0 \quad (10)$$

The governing equations are obtained substituting Eqs.(3 to 6) into equilibrium Eqs. (9, 10), which finally, reduce to the solution of

$$\nabla^2 u_z = 0 \text{ and } \nabla^2 \phi = 0 \quad (11)$$

where  $\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$  is the Laplacian operator.

Using Fourier cosine transform solution of Eq. (11) may be written as

$$u_z(x, y) = \frac{2}{\pi} \int_0^\infty [A_1(\alpha) \cosh(\alpha y) + A_2(\alpha) \sinh(\alpha y)] \cos(\alpha x) d\alpha + a_h y \quad (12)$$

and electric potential,  $\phi$  is given by

$$\phi(x, y) = \frac{2}{\pi} \int_0^\infty [B_1(\alpha) \cosh(\alpha y) + B_2(\alpha) \sinh(\alpha y)] \cos(\alpha x) d\alpha - b_h y \quad (13)$$

where  $A_i(\alpha), B_i(\alpha)$  are the arbitrary functions. These are determined using boundary conditions of the problem under investigation. And arbitrary constants  $a_h$  and  $b_h$  are obtained using conditions prescribed on the edges of strip.

Since the boundary condition are also prescribed on permittivity of the vacuum inside the crack, the constitutive equation for electric displacement components  $D_i^V, (i = x, y)$  reduce to

$$D_x^V = \epsilon_0 E_x \quad (14)$$

$$D_y^V = \epsilon_0 E_y \quad (15)$$

where  $\epsilon_0$  is the electrical permittivity of the vacuum and  $e_{15} = e_{31} = e_{33} = 0$ .

The governing equation for potential for potential  $\phi^V$  in vacuum reduce to

$$\nabla^2 \phi^V = 0 \quad (16)$$

The solution of which using Fourier transform technique and the condition  $D_y(x, 0) = D_y^V(x, 0)$  may be written as

$$\phi^V(x, y) = \frac{2}{\pi} \int_0^\infty C(\alpha) \sinh(\alpha y) \cos(\alpha x) d\alpha \text{ for } 0 \leq x < c \quad (17)$$

Opening mode electric displacement intensity factor, at the tip  $x = a$ , is defined as

$$K_I^D = \lim_{x \rightarrow a^+} [\sqrt{2\pi(x-a)} D_y(x, 0)] \quad (18)$$



In-plane open mode electrical displacement,  $D_y(x)$  is calculated using

$$D_y(x) = \frac{2}{e_{15}} \int_x^a M(x, \alpha) K_I^D(\alpha) d\alpha \quad (19)$$

where

$$M(x, \alpha) = \sqrt{\frac{\alpha}{\pi}} \frac{1}{\sqrt{\alpha^2 - x^2}} d\alpha \quad (20)$$

taken from ref. [10].

### 3 THE PROBLEM

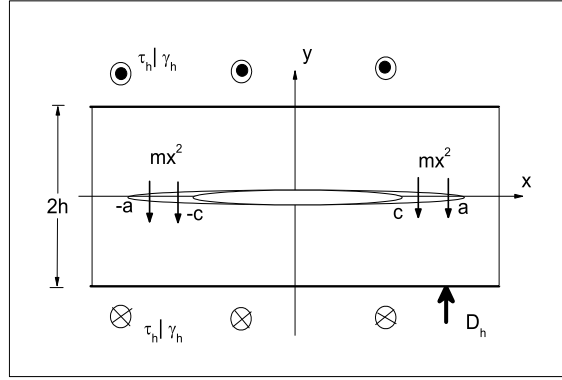


Figure 1: Schematic representation of the problem

An infinitely long narrow piezoceramic strip occupies the region  $-h \leq y \leq h$  and  $-\infty < x < \infty$  in  $xoy$ -plane as shown in Figure 1. The strip is assumed to be uniformly thick along  $z$ -direction to allow the anti-plane shear stress/strain state. The strip is poled along  $z$ -direction. The infinitely distant edges of the strip are stress/strain and charge free. The strip is cut along a hairline straight crack which occupies the interval  $y = 0$  and  $-c \leq x \leq c$  and oriented longitudinal to the edges of the strip. The rims of the crack are stress and charge free. The edges  $x = \pm h$  are prescribed uniform anti-plane shear stress  $\tau_{yz}(x, \pm h) = \tau_h$  or deformation  $\gamma_{yz}(x, \pm h) = \gamma_h$  together with in-plane normal electrical displacement  $D_y(x, \pm h) = D_h$ . Under the assumption that strip is electrically more brittle hence an electric singularity is encountered first. Consequently under small-scale electric polarization condition a strip-saturation zone develop ahead of each tip of the crack. Each of the saturation zone occupies the interval  $y = 0, c \leq x < a$  and  $-a < x \leq -c$ , respectively. To arrest further polarization the rims of the developed saturation zones are subjected to cohesive electrical displacement  $D_y = mx^2$ , where under small-scale electric saturation  $m = D_s/c^2$ ,  $D_s$  being the saturation limit electrical displacement. Consequently the crack is stopped from further opening.

## 4 MATHEMATICAL MODEL

A poled piezoceramic strip occupies the interval  $|y| = h$  and  $|x| < \infty$  in  $xoy$ -plane. The strip is cut along  $y = 0, |x| \leq c$ . Due to the symmetry in the problem only first quadrant region is considered. The conditions prescribed above may be mathematically get translated as

- (i) On finitely distant edges of the strip  $y = h, x \rightarrow \infty$ 
  - (a) *Case I* :  $\sigma_{yz}(x, h) = \tau_h, D_y(x, h) = D_h$
  - (b) *Case II* :  $\gamma_{yz}(x, h) = \gamma_h, D_y(x, h) = D_h$
- (ii)  $\phi(x, 0) = 0$ , for  $c \leq x < \infty$
- (iii)  $E_x(x, 0) = E_x^V(x, 0)$ , for  $0 \leq x < c$
- (iv)  $D_y(x, 0) = mx^2 H(x - c)$ , for  $0 \leq x < a$
- (v)  $u_z(x, 0) = 0$ , for  $a \leq x < \infty$

## 5 ANALYSIS AND SOLUTION

The general solution of the problem is written using Eqs. (12 and 13). Arbitrary functions and constants are determined using boundary condition (i to iv) as follows.

### 5.1 Determination of arbitrary constants $a_h$ and $b_h$

Substituting from Eqs. (12 and 13) into Eqs. (4, 6) and using each (i) and simplifying one obtains for

*Case I*:

$$a_h^I = \frac{\epsilon_{11}\tau_h - e_{15}D_h}{c_{44}\epsilon_{11} + e_{15}^2} \quad (21)$$

$$b_h^I = -\frac{e_{15}\tau_h - c_{44}D_h}{c_{44}\epsilon_{11} + e_{15}^2} \quad (22)$$

where superscript I denotes that the quantity refers to *Case I*.

And for *Case II* analogously using Eqs. (12, 13, 4, 6), boundary condition (i)(b) and calculating, following is obtained

$$a_h^{II} = \gamma_h \quad (23)$$

$$b_h^{II} = \frac{D_h - e_{15}\gamma_h}{\epsilon_{11}} \quad (24)$$

Superscript II denotes that the quantities refer to *Case II*.

### 5.2 Determination of arbitrary functions $A_i(\alpha)$ and $B_i(\alpha)$ , ( $i = 1, 2$ )

Remaining of the boundary condition (ii to v), using appropriate constitutive, gradient and Eqs. (12 and 13) yield a set of integral equations to enable to find

$A_i(\alpha)$  and  $B_i(\alpha)$  as follows:

Boundary condition (ii) together with Eq. (13) leads to integral equation

$$\int_0^\infty B_1(\alpha) \cos(\alpha x) d\alpha = 0; c \leq x < \infty \quad (25)$$

Boundary condition (iii) using Eqs. (13, 17, 8), yields the integral equation

$$\int_0^\infty \alpha B_1(\alpha) \sin(\alpha x) d\alpha = 0; 0 \leq x < c \quad (26)$$

Solving above pair Eqs. (25, 26) of dual integral and introducing  $B_1(\alpha)$  as

$$B_1(\alpha) = \frac{\pi c^2}{2} \int_0^1 \sqrt{\xi} \Psi_2(\xi) J_0(c\alpha\xi) d\xi \quad (27)$$

It is obtained that  $\Psi_2(\xi) = 0$  which implies that

$$B_1(\alpha) = 0 \quad (28)$$

Boundary condition (iv), Eqs. (12, 13, 6, 17, 15) can be simplified to yield

$$-\frac{2}{\pi} e_{15} \int_0^\infty \alpha A_1(\alpha) \tanh(\alpha h) \cos(\alpha x) d\alpha + d_0 = m x^2 H(x - c), 0 \leq x < a \quad (29)$$

$$d_0 = a_h^i + b_h^i \quad (30)$$

where  $i = I, II$ .

Boundary condition (v) together with Eq. (12) gives

$$\int_0^\infty A_1(\alpha) \cos(\alpha x) d\alpha = 0; a \leq x < \infty \quad (31)$$

Introducing for the convenience of computations

$$A_1(\alpha) = \frac{\pi a^2}{2} \int_0^1 \sqrt{\xi} \Psi_1(\xi) J_0(a\alpha\xi) d\xi \quad (32)$$

and solving the pair of dual integral Eqs. (29, 31), one finally obtains after computations a Fredholm integral equation of second kind for determining  $\Psi_1(\xi)$  from

$$\begin{aligned} \Psi_1(\xi) + \int_0^1 K(\xi, \eta) \Psi_1(\eta) d\eta = \\ \begin{cases} \frac{D_h \xi^{1/2}}{e_{15}}, & \xi < \frac{c}{a} \\ \frac{D_h \xi^{1/2}}{e_{15}} - \frac{m a^2 \xi^{5/2}}{2 e_{15}} \left( 1 - \frac{2}{\pi} \arcsin\left(\frac{c/a}{\xi}\right) + \frac{1}{\pi} \sin\left(2 \arcsin\left(\frac{c/a}{\xi}\right)\right) \right), & \frac{c}{a} < \xi < 1 \end{cases} \end{aligned} \quad (33)$$

$A_1(\alpha)$  is now determined using Eqs. (33 and 32).

where kernel  $K(\xi, \eta)$

$$K(\xi, \eta) = \sqrt{\xi \eta} \int_0^\infty \alpha \tanh\left(\frac{\alpha h}{a} - 1\right) J_0(\alpha \eta) J_0(\alpha \xi) d\alpha \quad (34)$$

Equation (33) in turn is solved numerically using *MATHEMATICA*.

## 6 APPLICATIONS

Results obtained are applied to calculate open mode electrical displacement intensity factor  $K_I^D(a)$ , saturation zone length, in-plane electric crack opening displacement, crack growth rate of electrical crack opening displacement.

### 6.1 Open-mode electrical displacement intensity factor $K_I^D(a)$

Substituting in formula (18) from Eqs. (6, 12, 13, 28, 32, 33 and 34) and simplifying one obtains the  $K_I^D(x)$  at  $x = a$ .

$$K_I^D(a) = e_{15} \sqrt{\pi a} \Psi_1(1) \quad (35)$$

### 6.2 Saturation zone length

Saturation zone is obtained using the hypothesis that electrical polarization vanishes at  $x = a$ , based on Dugdale hypothesis. Consequently Eq. (35) yields  $\Psi_1(1) = 0$ .

Equating  $\Psi_1(1)$  using Eq. (33 and 34) gives a transcendental equation to determine  $a$  from

$$a^2 \cos^{-1} \left( \frac{c}{a} \right) + c^2 \sqrt{\frac{a^2}{c^2} - 1} = \pi \left[ \frac{d_0 - e_{15} T(h/a)}{m} \right] \quad (36)$$

where

$$T(h/a) = \int_{\eta=0}^1 K(1, \eta) \Psi_1(\eta) d\eta \quad (37)$$

## 7 CRACK OPENING IN-PLANE ELECTRICAL DISPLACEMENT

Crack opening in-plane electrical displacement,  $D_y(x)$ , is obtained superimposing the  $[D_y]_{D_h}$  due to electrical displacement,  $D_h$ , prescribed on the edges of the strip and the normal cohesive in-plane electric displacement  $[D_y]_{mx^2}$  prescribed on the rims of the developed saturation zone.

Using formulae (19, 20) and substituting  $[K_I^D(\alpha)]_{D_h} = d_0 \sqrt{\pi \alpha}$  and integrating, one obtains

$$[D_y(x)]_{D_h} = \frac{2d_0}{e_{15}} \sqrt{a^2 - x^2} \quad (38)$$

Similarly  $[D_y(x)]_{mx^2}$  is obtained using formulae (19, 20) and  $[K_I^D(\alpha)]_{mx^2} = \frac{m}{\sqrt{\pi}} \alpha^{5/2} \left[ \cos^{-1} \frac{c}{a} + \frac{c \sqrt{\alpha^2 - x^2}}{\alpha^2} \right]$  and integrating

$$[D_y(x)]_{mx^2} = \frac{2m}{\pi e_{15}} \int_{\alpha=x}^a \frac{\alpha^3}{\sqrt{\alpha^2 - x^2}} \left[ \frac{\pi}{2} - \sin^{-1} \frac{c}{\alpha} + \frac{1}{2} \sin \left( 2 \sin^{-1} \left( \frac{c}{\alpha} \right) \right) \right] d\alpha \quad (39)$$

Consequently the crack opening displacement of crack face is obtained using

$$D_y(x) = [D_y(x)]_{D_h} - [D_y(x)]_{mx^2} \quad (40)$$

Total crack opening displacement is obtained by  $2D_y(x)$ .

## 8 ELECTRIC DISPLACEMENT GROWTH RATE

Under the assumption that cyclic condition could be obtained from the prescribed monotonic condition. These could be obtained by substitution

$$D_h \rightarrow \frac{\Delta D_h}{2}, D_0 \rightarrow \frac{\Delta D_0}{2}, D_s \rightarrow D_{sc}$$

where  $\Delta$  in front of a quantity denotes that quantity is under cyclic conditions.  $D_0$  is the electrical displacement in absence of mechanical load. The accumulated crack opening displacement is defined as

$$\Sigma|\Delta D_y(x)| = \frac{2}{\delta c} \int_c^a \Delta D_y(x) dx, \quad (41)$$

where  $\delta c = \frac{dc}{dN}$  is the crack growth rate, where  $N$  is number of loading cycles which is equal to  $(a-c)/\delta c$  for the crack to advance a distance  $(a-c)$  for applied electric displacement range.

The fatigue crack propagation [11] is assumed to start when the accumulated electric displacement  $D_T$  of a certain point equals a critical value,  $D_c$  and gave a crack growth rate criteria

$$D_T = \Sigma|\Delta D_y(x)| = D_c \quad (42)$$

This together with Eqs. (41 and 42) gives

$$\begin{aligned} \frac{dc}{dN} &= \frac{2}{D_c} \int_c^a \Delta D_y(x) dx, \\ &= \frac{8mc^4}{\pi e_{15} D_c} \left[ \frac{5}{12} - \frac{5}{12} \left(\frac{a}{c}\right)^2 - \frac{1}{6} \left(\frac{a}{c}\right)^2 \sqrt{\frac{a^2}{c^2} - 1} \left\{ \cos^{-1} \left(\frac{c}{a}\right) \right\} \right. \\ &\quad \left. + \frac{2}{3} \sqrt{\frac{a^2}{c^2} - 1} \left\{ \cos^{-1} \left(\frac{c}{a}\right) \right\} + \frac{1}{4} \left(\frac{a}{c}\right)^4 \left\{ \cos^{-1} \left(\frac{c}{a}\right) \right\}^2 + \frac{2}{3} \log \left(\frac{c}{a}\right) \right] \end{aligned} \quad (43)$$

which to first approximation may be written as

$$\frac{dc}{dN} = \frac{\pi \{ \Delta K_I^D(a) \}^4}{192 e_{15} \gamma D_{sc}^2} \quad (44)$$

where  $\gamma = D_c D_{sc}$ , is the effective surface energy for crack growth and

$$\Delta K_I^D = [\Delta D_h - 2e_{15} T(h/a)] \sqrt{\pi c} \quad (45)$$

## 9 CASE STUDY

Results obtained are applied to investigate the crack arrest for  $\text{BaTiO}_3$ , PZT-4 and PZT-5H strips. Material constants are taken from [6] and listed in Table.I.

Table I: Material constants for piezoelectric ceramics.

Material Constants	PZT-4	PZT-5H	$\text{BaTiO}_3$
$c_{44}(10^{10} \text{ N/m}^2)$	2.56	2.3	4.3
$e_{15}(\text{ C/m}^2)$	12.7	17	11.6
$\epsilon_{11}(10^{-10} \text{ C/Vm})$	64.6	150.4	112

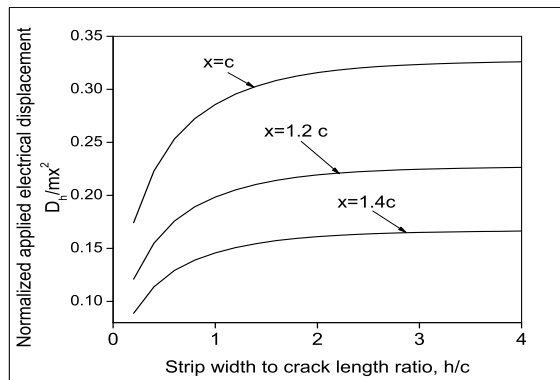


Figure 2: Variation of applied electrical displacement versus strip width to half-crack length ratio

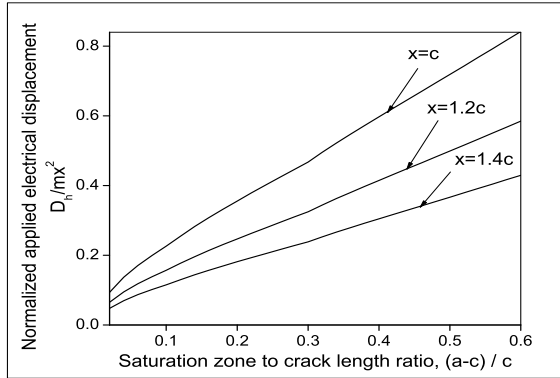


Figure 3: Variation of applied electrical displacement versus saturation zone length to half-crack length ratio

Figure 2. depicts the variation of variable load ratio  $D_h/mx^2$  for small scale

yielding, with respect to strip width to crack length ratio  $h/c$ . For narrow strip a non-linear parabolic variation is observed for  $D_h/mx^2$ , as expected. But as the thickness of the strip is increased for a fixed crack length then more  $D_h/mx^2$  required to arrest the crack before the requirement settles for a uniform constant value. It is also noted as  $x/c$  ratio is increased then less load ratio is required to arrest the crack growth.

Variation of  $D_h/mx^2$  vis-a-vis saturation to half-crack length,  $(a - c)/c$ , is plotted in Figure 3. It is seen as the saturation zone is increased, for a fixed crack length, more  $D_h/mx^2$  is needed, as expected. It is also to be noted that variation remains same for both the cases.

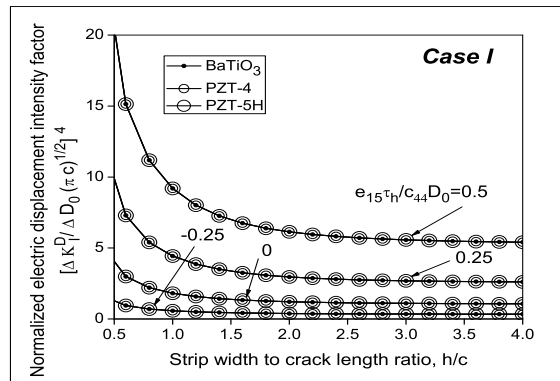


Figure 4: Variation of normalized electric crack growth rate versus strip width to half-crack length ratio for *Case I*

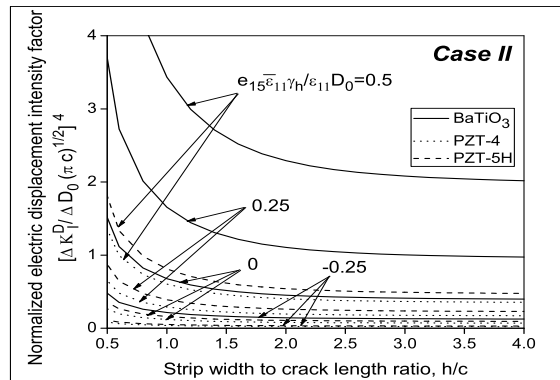


Figure 5: Variation of normalized electric crack growth rate versus strip width to half-crack length ratio for *Case II*

For *Case I* crack growth variations versus strip width to crack length ratio is depicted in Figure 4. The variation first shows a steep parabolic decrease

which stabilizes uniformly to a lower value as strip width is increased. The crack growth is also independent of ceramic properties. It is seen as the ratio value  $p = e_{15}\tau_h/c_{44}D_0$  is increased as 0.5(-0.25)-0.25, the crack growth almost becomes negligible.

Same variation for *Case II* is plotted in Figure 5. The crack growth for this case, is almost four times smaller than that for *Case I*. The variation is material properties dependent. PZT-4 ceramic strip has the least crack growth rate closely followed by PZT-5H strip. For lower values of the ratio  $q = e_{15}\bar{\epsilon}_{11}\gamma_h/\epsilon_{11}D_0 = -0.25$  the crack growth rates for both the ceramics PZT-4 and PZT-5H are negligibly small. Comparatively although BaTiO<sub>3</sub> has higher crack growth rate but it also drops down substantial as the ratio value  $q$  is reduced from 0.5 to -0.25.

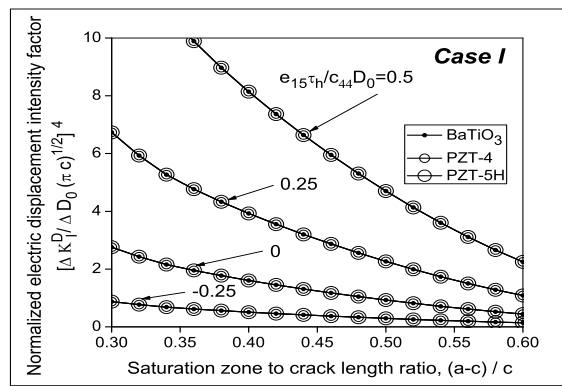


Figure 6: Variation of normalized electric crack growth rate versus saturation zone to half-crack length ratio for *Case I*

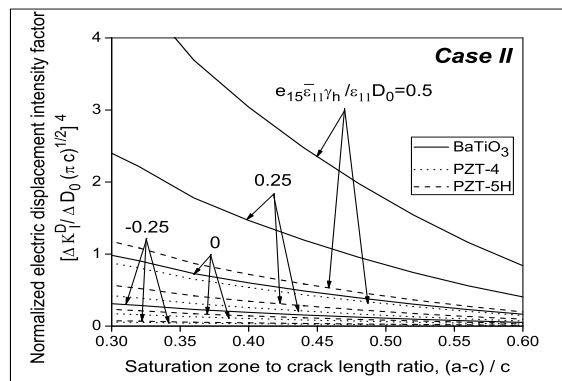


Figure 7: Variation of normalized electric crack growth rate versus saturation zone to half-crack length ratio for *Case II*



For *Case I* electric crack growth rate variation vis-a-vis saturation zone to half-crack length,  $(a - c)/c$ , is plotted in Figure 6. The crack growth rate is independent of ceramic properties and show a continuous decrease as  $(a - c)/c$  ratio is increased. It is also noted as the value of the ratio  $p$  is reduced from 0.5 to -0.25 the crack growth rate reduces further tending to almost zero.

Figure 7. depicts the variation of crack growth rate with respect to ratio  $(a - c)/c$  for *Case II*. For *Case II* the crack growth reduces to more than half than that for *Case I*. The crack growth rate is least for PZT-4 strip which almost zero as  $q$  ratio is reduced from 0.5 to -0.25. Although BaTiO<sub>3</sub> strip has a higher crack growth rate but it continuously reduces to almost become zero. As the ratio value  $q$  is reduced from 0.5 to -0.25 it drops down further.

## 10 CONCLUSION

A generalized strip-saturation model is proposed for a longitudinally cracked poled piezoceramic strip. The analytic expressions are obtained for electrical crack opening displacement and crack growth rate. Electrical displacement required to be prescribed for saturation zone to arrest crack are also obtained in closed form expressions. Case study has been presented for BaTiO<sub>3</sub>, PZT-4 and PZT-5H strips which shows that the model proposed is capable of crack arrest.

## Acknowledgement

Authors are grateful to **Prof. R. D. Bhargava** {Senior Professor and Head (retd.), Indian Institute of Technology Bombay, Mumbai} for his valuable advice and continuous encouragement. The financial support by Council of Scientific and Industrial Research, New Delhi, during the course of this work is thankfully acknowledged.

## References

- [1] B. L. Wang and X. H. Zhang, Fracture prediction for piezoelectric ceramics based on the electric field saturation concept, *Mech. Res. Commun.*, 32, 411-419 (2005).
- [2] B. L. Wang and Y. W. Mai, A cracked piezoelectric material strip under transient thermal loading, *Trans. ASME, J. Appl. Mech.*, 69, 539-546 (2002).
- [3] C. Li and G. J. Weng, Antiplane crack problem in functionally graded piezoelectric materials, *Trans. ASME, J. Appl. Mech.*, 69, 481-488 (2002).
- [4] C. Q. Ru, Effect of electrical polarization saturation on stress intensity factors in a piezoelectric ceramic, *Int. J. Solids Struct.*, 36, 869-883 (1999).

- [5] C. Q. Ru and X. Mao, Conducting cracks in a piezoelectric ceramic of limited electrical polarization, *J. Mech. Phys. Solids*, 47, 2125-2146 (1999).
- [6] F. Narita and Y. Shindo, Fatigue crack propagation in a piezoelectric ceramic strip subjected to mode III loading, *Acta Mech.*, 137, 55-63 (1999).
- [7] H. Gao, T. -Y. Zhang and P. Tong, Local and global energy release rates for an electrically yielded crack in a piezoelectric ceramic, *J. Mech. Phys. Solids*, 45, 491-510 (1997).
- [8] H. Gao and D. M. Barnett, An invariance property of local energy release rate in a strip saturation model of piezoelectric fracture, *Int. J. Fract.*, 79, 25-29 (1996).
- [9] S. Li, On saturation strip model of a permeable crack in a piezoelectric ceramic, *Acta Mech.*, 165, 47-71 (2003).
- [10] S. R. Daniewicz, A closed-form small-scale yielding collinear strip yield model for strain hardening material, *Eng. Fract. Mech.*, 49, 95-103 (1994).
- [11] W. -R. Chen and L. M. Keer, Fatigue crack growth in mixed mode loading, *Trans. ASME, J. Eng. Mater. Technol.*, 113, 222-227 (1991).
- [12] Y. Shindo, F. Narita and K. Tanaka, Electroelastic intensification near anti-plane shear crack in orthotropic piezoelectric ceramic strip, *Theor. Appl. Fract. Mech.*, 25, 65-71 (1996).
- [13] Y. Shindo, K. Tanaka and F. Narita, Singular stress and electric fields of a piezoelectric ceramic strip with a finite crack under longitudinal shear, *Acta Mech.*, 120, 31-45 (1997).

# On the uniqueness of the Fourier projection in $L_p$ spaces

Boris Shekhtman and Lesław Skrzypek

Department of Mathematics and Statistics  
University of South Florida  
4202 E Fowler Ave  
Tampa FL 33620, USA

`boris@math.usf.edu` (*Boris Shekhtman*)  
`skrzypek@math.usf.edu` (*Lesław Skrzypek*)

March 14, 2009

## Abstract

Let  $\pi_N$  stand for the space of trigonometric polynomials of degree  $N$  and  $F : L_p[0, 2\pi] \rightarrow \pi_N$  denote the Fourier projection onto  $\pi_N$ . It is well-known that, for all  $1 \leq p \leq \infty$ ,  $F$  is the projection onto  $\pi_N$  with the minimal norm. It is an open problem whether for every  $1 < p < \infty$ ,  $p \neq 2$ ,  $F$  is a unique minimal projection onto  $\pi_N$ . The aim of this paper is to prove the unique minimality of  $F$  in  $L_p[0, 2\pi]$  spaces for  $p$  sufficiently closed to 1.

## 1 Introduction

Let  $\pi_N = \text{span}\{1, \sin(x), \cos(x), \dots, \sin(Nx), \cos(Nx)\}$  stand for the space of trigonometric polynomials of degree  $N$ . The Fourier projection  $F : L_p[0, 2\pi] \rightarrow \pi_N$ ,  $1 \leq p \leq \infty$  is defined by

$$F(f)(x) = \frac{1}{2\pi} \int_0^{2\pi} f(t) dt + \sum_{k=1}^N \left( \frac{1}{\pi} \int_0^{2\pi} f(t) \sin(kt) dt \right) \sin(kx) \\ + \sum_{k=1}^N \left( \frac{1}{\pi} \int_0^{2\pi} f(t) \cos(kt) dt \right) \cos(kx). \quad (1)$$

One can see that  $F(f) = f * D_N$ , that is

$$F(f)(x) = \frac{1}{2\pi} \int_0^{2\pi} f(t) D_N(x-t) dt, \quad (2)$$

where  $D_N$  is a Dirichlet kernel

$$D_N(x) = 1 + 2 \cos(x) + \dots + 2 \cos(Nx). \quad (3)$$

Using the Dirichlet kernel we can compute the norm of the Fourier projection in  $L_1[0, 2\pi]$

$$\|F\|_1 = \frac{1}{2\pi} \int_0^{2\pi} |D_N(x)| dx = L_N, \quad (4)$$

where  $L_N$  are famous Lebesgue's constants. There are two exact formulas for Lebesgue's constants, the Fejer formula

$$L_N = \frac{1}{2N+1} + \sum_{k=1}^N \frac{2}{k\pi} \tan\left(\frac{k\pi}{2N+1}\right) \quad (5)$$

and the Szegö formula

$$L_N = \frac{16}{\pi^2} \sum_{k=1}^{\infty} \left(1 + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{2(2N+1)k-1}\right) \frac{1}{4k^2-1}. \quad (6)$$

Additionally, the exact asymptotic behavior of  $L_N$  is known to be  $L_N \sim \frac{4}{\pi^2} \ln N$ . Consider the following isometries  $I_s : L_p[0, 2\pi] \rightarrow L_p[0, 2\pi]$ ,  $s \in [0, 2\pi)$ ,

$$I_s(f)(x) = f(x+s), \quad (7)$$

where the addition of  $x+s$  is considered to be *modulo*  $[0, 2\pi]$ . One can easily see that  $I_s(\pi_N) = \pi_N$ , and that the Fourier projection is the only projection that is commuting with  $\{I_s, s \in [0, 2\pi)\}$

$$F \circ I_s = I_s \circ F. \quad (8)$$

Following this observation, we get the Berman representation for the Fourier projection [1],

$$F = \frac{1}{2\pi} \int_0^{2\pi} (I_s)^{-1} \circ P \circ I_s ds, \quad (9)$$

where  $P$  is any projection  $L_p[0, 2\pi] \rightarrow \pi_N$ .

From that we obtain minimality of Fourier projection in  $L_p[0, 2\pi]$  spaces. Minimality of Fourier projection in  $C[0, 2\pi]$  has been proved earlier in 1948 by Lozinski in [6]. Whether the Fourier projection is the unique minimal or not has proved to be a challenging problem. In 1969, Cheney et al [2], have proved that the Fourier projection is a unique minimal in  $C[0, 2\pi]$  and  $L_\infty[0, 2\pi]$ . That same year, Lambert [4], proved the same for  $L_1[0, 2\pi]$ . The case of  $L_p[0, 2\pi]$  for  $1 < p < \infty, p \neq 2$  is still open. In our previous paper, [9], we proved that for  $N = 1$  the Fourier projection  $F : L_p[0, 2\pi] \rightarrow \pi_1$  is unique minimal.  $L_p[0, 2\pi]$  for  $1 < p < \infty, p \neq 2$  are smooth and uniformly convex spaces. It is worth mentioning that for any  $d \geq 3$  we can find a finite dimensional subspace  $V$ ,  $\dim V = d$  such that the minimal projection from  $L_p[0, 2\pi]$  onto  $V$  is not a

unique minimal [11]. For  $d = 1, 2$  all minimal projections have to be unique [10]. Uniqueness of related Rademacher projections in  $L_p[0, 2\pi]$ ,  $1 < p < \infty$  has been proved in [5]. In this paper we prove that Fourier projection is unique minimal in  $L_p[0, 2\pi]$  for  $p$  sufficiently close to 1.

**Definition 1.1** *Let  $L : X \rightarrow Y$  be a linear operator. Functional  $g \in S(Y^*)$  is called a norming functional for  $L$  if*

$$\|g \circ L\| = \|L\|. \quad (10)$$

*Point  $f \in S(X)$  is called a norming point for  $L$  if*

$$\|L(x)\| = \|L\|. \quad (11)$$

*A pair  $(g, f) \in S(Y^*) \times S(X)$  is called a norming pair for  $L$  if*

$$g(Lf) = \|L\|. \quad (12)$$

If  $P$  is a projection from  $X$  onto a finite-dimensional subspace  $Y$ , then (since  $P$  is a compact operator) it has a norming functional (see Theorem III.2.1 [7]). If  $X$  is reflexive, then any functional attains its norm. Therefore, there is a norming pair for  $P$ . If  $X$  is not reflexive then in general, it is not true. For example, the Fourier projection does not attain its norm in  $C[0, 2\pi]$ . But any functional attains its norm in  $X^{**}$ , hence we can always find a norming pair for  $P$  (extending  $P$  to  $P^{**}$ ) in  $S(X^*) \times S(X^{**})$ .

The properties of norming functionals play important roles in proving the uniqueness of a minimal projection. The following theorem is a special case of the result proved in [9].

**Theorem 1.2 (Theorem 3.2 [9])** *Let  $F : L_p[0, 2\pi] \rightarrow \pi_N$  be the Fourier projection. Assume that the set of norming functionals for  $F$  is total on  $\pi_N$  (i.e., if  $v \in \pi_N$ , and  $f(v) = 0$  for any norming functional  $f \in L_q[0, 2\pi]$  for  $F$ , then  $v = 0$ .) Then the Fourier projection  $F$  is the unique minimal projection.*

The main result of this paper is the following:

**Theorem 1.3** *Fix  $N$  and consider the Fourier projection  $F : L_p[0, 2\pi] \rightarrow \pi_N$ . Then there is an  $\epsilon > 0$  such that for every  $p \in (1, 1 + \epsilon)$  the Fourier projection is the unique minimal projection from  $L_p[0, 2\pi]$  onto  $\pi_N$ .*

## 2 Results

We begin with some preliminaries.

**Lemma 2.1 (Ruess-Stegall [8])** *Let  $X$  be a Banach space and  $V$  be the finite dimensional Banach space. Then*

$$L(X, V) = L_{w*}(X^{**}, V), \quad (13)$$

where " $=$ " means isometrically isomorphic and  $L_{w*}(X^{**}, V)$  stands for all weak\* continuous operators. Additionally, this isometry is given by

$$i : L \mapsto L^{**}. \quad (14)$$

Since  $\frac{1}{2\pi} dm$  is a probabilistic measure, for every  $1 \leq r \leq s \leq \infty$  we have

$$L_\infty[0, 2\pi] \subset L_s[0, 2\pi] \subset L_r[0, 2\pi] \subset L_1[0, 2\pi] \quad (15)$$

**Theorem 2.2** Fix  $N$  and consider the Fourier projection  $F : L_p[0, 2\pi] \rightarrow \pi_N$ . Take any  $p_n \rightarrow 1$  and let  $g_n \in S(L_{q_n}[0, 2\pi])$  be any norming functional for  $F$  in  $L_{p_n}[0, 2\pi]$  (as usual  $p_n$  and  $q_n$  are conjugates to each other  $\frac{1}{p_n} + \frac{1}{q_n} = 1$ ). Then we can find a subsequence of  $g_n$  that converges weakly in  $L_1[0, 2\pi]$  to some function  $g$ ,  $|g| = 1$  a.e. such that  $g$  is a norming functional for  $F$  in  $L_1[0, 2\pi]$ .

**Proof.** Take  $f_n \in S(L_{p_n}[0, 2\pi])$  such that  $(g_n, f_n)$  is a norming pair for  $F$  in  $L_{p_n}[0, 2\pi]$ , that is

$$g_n(Ff_n) = \|F\|_{p_n}. \quad (16)$$

Since  $\|F\|_{p_n} \rightarrow \|F\|_1$ , we also have

$$\|F(f_n)\|_{p_n} \rightarrow \|F\|_1. \quad (17)$$

Observe that  $F(f_n) \in \pi_N \subset L_1[0, 2\pi]$ , and  $\|F(f_n)\|_1 \leq \|F(f_n)\|_{p_n} = \|F\|_{p_n}$ . Therefore  $F(f_n)$  is bounded in  $\pi_N$  with  $L_1$  norm. Since  $\pi_N$  is finite dimensional  $F(f_n)$  is bounded in any other norm. As a result going into subsequence if necessary we can assume that

$$F(f_n) \rightarrow w \in \pi_N. \quad (18)$$

Observe that  $w \neq 0$  (otherwise  $F(f_n) \rightarrow 0$  in  $L_\infty$  norm, and that for a big enough  $n$  would imply  $\|F\|_{p_n} = \|F(f_n)\|_{p_n} < 1/2$ ). Therefore  $w$  has only a finite number of zeroes. Using (16) for  $n$  big enough we have

$$|g_n(t)| = \frac{|(Ff_n)(t)|^{p_n/q_n}}{(\|F\|_{p_n})^{p_n/q_n}}, \quad (19)$$

for all  $t$  except possibly finite many. Since  $p_n/q_n = p_n - 1 \rightarrow 0$  and  $F(f_n)(t) \rightarrow w(t)$  (see (18)) using the continuity of the function  $h(x, y) = x^y$  at points  $(a, 0)$  where  $a > 0$  we get  $|(Ff_n)(t)|^{p_n/q_n} \rightarrow 1$  and  $(\|F\|_{p_n})^{p_n/q_n} \rightarrow 1$ . Therefore

$$|g_n(t)| \rightarrow 1, \quad (20)$$

for all  $t$  except possibly finite many. As a result

$$g_n(t) \rightarrow g(t), \quad (21)$$

almost everywhere,  $|g(t)| = 1$  and therefore  $g \in S(L_\infty[0, 2\pi])$ . Observe that both  $F(f_n)$  and  $w$  are continuous functions. Using (18) there is a constant  $M$

such that  $|F(f_n)(t)| < M$ . By (19) there is a constant  $K$  such that  $|g_n(t)| < K$  for all  $t$  except finite many points. That means that  $g_n \in L_\infty[0, 2\pi]$  and by Lebesgue's dominated convergence theorem:

$$\int_0^{2\pi} g_n(t)h(t) dm \rightarrow \int_0^{2\pi} g(t)h(t) dm, \quad (22)$$

for every  $h \in L_1[0, 2\pi]$ .

Now we need to show that this  $g$  is a norming functional for  $F$  in  $L_1[0, 2\pi]$ . Consider  $F^{**} : L_1^{**} \rightarrow \pi_N$ . Observe that

$$\|f_n\|_1 \leq \|f_n\|_{p_n} = 1. \quad (23)$$

Therefore  $f_n \in L_1[0, 2\pi]$  and using canonical embedding of  $X$  in  $X^{**}$  we have  $f_n \in L_1^{**}[0, 2\pi]$  and  $\|f_n\|_{L_1^{**}} \leq 1$ . As a result  $f_n$  has a weak\* convergent subsequence in  $L_1^{**}$ . Passing to subsequence, if necessary, we may assume that

$$f_n \xrightarrow{w*} f, \quad (24)$$

weak\* in  $L_1^{**}[0, 2\pi]$  and  $\|f\|_{L_1^{**}} \leq 1$ . By Lemma 2.1, since  $F^{**}(f_n), F^{**}(f) \in \pi_N$ , we have

$$F^{**}(f_n) \rightarrow F^{**}(f), \quad (25)$$

in norm topology. By (22) we have

$$g_n(h) \rightarrow g(h), \quad (26)$$

for every  $h \in \pi_N$ . Putting the above two facts together we get

$$\begin{aligned} |g_n(F^{**}f_n) - g(F^{**}f)| &= |g_n(F^{**}f_n - F^{**}f) + (g_n - g)(F^{**}f)| \\ &\leq \|F^{**}f_n - F^{**}f\| + |(g_n - g)(F^{**}f)| \rightarrow 0. \end{aligned} \quad (27)$$

But  $g_n(F^{**}f_n) = g_n(Ff_n) = \|F\|_{p_n} \rightarrow \|F\|_1 = \|F\|_{L_1^{**}}$ . Therefore

$$g(F^{**}f) = \|F\|_{L_1^{**}}. \quad (28)$$

And since  $\|f\|_{L_1^{**}} \leq 1$  we have  $\|g \circ F^{**}\| = \|F\|_{L_1^{**}}$  and as a result

$$\|g \circ F\| = \|F\|_1, \quad (29)$$

so  $g$  is a norming functional for  $F$  in  $L_1[0, 2\pi]$ . ■

**Lemma 2.3** *Every norming functional  $g \in S(L_\infty[0, 2\pi])$  for the Fourier projection  $F$  in  $L_1[0, 2\pi]$  has to be of the form*

$$g(t) = \text{sign } D_N(t + s), \quad (30)$$

for some  $s \in [0, 2\pi]$  (here the addition is considered to be modulo  $[0, 2\pi]$ ).

**Proof.** Let  $g \in S(L_\infty[0, 2\pi])$  be a norming functional for Fourier projection  $F : L_1[0, 2\pi] \rightarrow \pi_N$ . Denote  $dt = \frac{1}{2\pi} dm(t)$  and  $dx = \frac{1}{2\pi} dm(x)$ . For any  $t$  we have

$$\begin{aligned} \left| \int_0^{2\pi} g(x) D_N(x-t) dx \right| &\leq \int_0^{2\pi} |g(x) D_N(x-t)| dx \\ &= \int_0^{2\pi} |D_N(x-t)| dx = \int_0^{2\pi} |D_N(x)| dx. \end{aligned} \quad (31)$$

Since  $D_N(x) = 1 + 2 \cos(x) + \dots + 2 \cos(Nx)$ , the function

$$h(t) = \int_0^{2\pi} g(x) D_N(x-t) dx \quad (32)$$

is continuous on  $[0, 2\pi]$ . We will show that

$$\max_{t \in [0, 2\pi]} \left| \int_0^{2\pi} g(x) D_N(x-t) dx \right| = \int_0^{2\pi} |D_N(x)| dx. \quad (33)$$

Assume for the contrary that there is  $\delta > 0$

$$\max_{t \in [0, 2\pi]} \left| \int_0^{2\pi} g(x) D_N(x-t) dx \right| = \left( \int_0^{2\pi} |D_N(x)| dx \right) - \delta \quad (34)$$

Take any  $f \in S(L_1[0, 2\pi])$ . Using Fubini's Theorem we would get

$$\begin{aligned} |g(Ff)| &= \left| \int_0^{2\pi} \left( \int_0^{2\pi} f(t) g(x) D_N(x-t) dt \right) dx \right| \\ &= \left| \int_0^{2\pi} \left( \int_0^{2\pi} f(t) g(x) D_N(x-t) dx \right) dt \right| \\ &= \left| \int_0^{2\pi} f(t) \left( \int_0^{2\pi} g(x) D_N(x-t) dx \right) dt \right| \\ &\leq \int_0^{2\pi} |f(t)| \left| \int_0^{2\pi} g(x) D_N(x-t) dx \right| dt \\ &\leq \int_0^{2\pi} |f(t)| \left( \left( \int_0^{2\pi} |D_N(x)| dx \right) - \delta \right) dt \\ &= \left( \left( \int_0^{2\pi} |D_N(x)| dx \right) - \delta \right) \left( \int_0^{2\pi} |f(t)| dt \right) \\ &= \left( \int_0^{2\pi} |D_N(x)| dx \right) - \delta = \|F\|_1 - \delta. \end{aligned} \quad (35)$$

That implies that  $g$  is not a norming functional for  $F$ , a contrary. Therefore (33) holds and, as a result of (31), there is  $t_0$  such that

$$\left| \int_0^{2\pi} g(x) D_N(x-t_0) dx \right| = \int_0^{2\pi} |g(x) D_N(x-t_0)| dx, \quad (36)$$

and as a result  $g(x) = \text{sign } D_N(x-t_0)$ . ■



**Theorem 2.4** Fix  $N$  and consider the Fourier projection  $F : L_p[0, 2\pi] \rightarrow \pi_N$ . Then there is an  $\epsilon > 0$  such that for every  $p \in (1, 1 + \epsilon)$  we can find a norming functional  $h \in S(L_q[0, 2\pi])$  for  $F$  in  $L_p[0, 2\pi]$  such that

$$\left(\int_0^{2\pi} h(t) \sin(kt) dt\right)^2 + \left(\int_0^{2\pi} h(t) \cos(kt) dt\right)^2 \neq 0, \text{ for } k = 0, 1, \dots, N. \quad (37)$$

**Proof.** Assume for the contrary that there is a sequence  $p_n \rightarrow 1$  and a sequence of norming functionals  $g_n \in S(L_{q_n}[0, 2\pi])$  for  $F$  in  $L_{p_n}[0, 2\pi]$  such that

$$\int_0^{2\pi} g_n(t) \sin(kt) dt = 0 \quad \text{and} \quad \int_0^{2\pi} g_n(t) \cos(kt) dt = 0, \quad (38)$$

for some  $k \in \{0, 1, \dots, N\}$ . Using Theorem 2.2 we get

$$\int_0^{2\pi} g(t) \sin(kt) dt = 0 \quad \text{and} \quad \int_0^{2\pi} g(t) \cos(kt) dt = 0 \quad (39)$$

and  $g$  is a norming functional for  $F$  in  $L_1[0, 2\pi]$ . But Lemma 2.3 gives  $g(t) = \text{sign } D_n(x+s)$ . Using Fejer formula [3], we know the first  $2N+1$  terms in Fourier expansion of  $\text{sign } D_n(t)$  :

$$\text{sign } D_n(t) = \frac{1}{2N+1} + \sum_{m=1}^N \frac{2}{m\pi} \tan \frac{m\pi}{2N+1} \cos(mt) + \dots \quad (40)$$

From the above we can easily see now that (39) cannot occur. ■

**Theorem 2.5** Fix  $N$  and consider the Fourier projection  $F : L_p[0, 2\pi] \rightarrow \pi_N$ . Then there is an  $\epsilon > 0$  such that for every  $p \in (1, 1 + \epsilon)$  the Fourier projection is the unique minimal projection from  $L_p[0, 2\pi]$  onto  $\pi_N$ .

**Proof.** Take  $\epsilon$  and functional  $h$  from Theorem 2.4. Observe that, by (8), if  $h(t)$  is a norming functional then  $h_s(t) = h(t+s)$  is also a norming functional. We will show that the set of norming functionals  $\{h_s, s \in [0, 2\pi]\}$  is total over  $\pi_N$ . That is if  $w(x) \in \pi_N$  and for every  $s \in [0, 2\pi]$  :

$$\frac{1}{2\pi} \int_0^{2\pi} h(t+s)w(t) dt = 0, \quad (41)$$

then  $w = 0$ . Take

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_0^{2\pi} h(t) dt \\ a_k &= \frac{1}{\pi} \int_0^{2\pi} h(t) \sin(kt) dt \\ b_k &= \frac{1}{\pi} \int_0^{2\pi} h(t) \cos(kt) dt. \end{aligned} \quad (42)$$

By Theorem 2.4  $a_0 \neq 0$  and  $a_k, b_k \neq 0$  for  $k = 1, \dots, N$ . It is easy to see that

$$\begin{aligned} \frac{1}{2\pi} \int_0^{2\pi} h(t+s) dt &= a_0 \\ \frac{1}{\pi} \int_0^{2\pi} h(t+s) \sin(kt) dt &= a_k \cos(ks) - b_k \sin(ks) \\ \frac{1}{\pi} \int_0^{2\pi} h(t+s) \cos(kt) dt &= a_k \sin(ks) + b_k \cos(ks). \end{aligned} \quad (43)$$

Take  $w = c_0 + \sum_{k=1}^N (c_k \sin(kt) + d_k \cos(kt))$ . By (43), the equation (41) would imply

$$a_0 c_0 + \sum_{k=1}^N [c_k (a_k \cos(ks) - b_k \sin(ks)) + d_k (a_k \sin(ks) + b_k \cos(ks))] = 0, \quad (44)$$

for every  $s \in [0, 2\pi]$ . As a result

$$a_0 c_0 + \sum_{k=1}^N [(d_k a_k - c_k b_k) \sin(ks) + (c_k a_k + d_k b_k) \cos(ks)] = 0, \quad (45)$$

for every  $s \in [0, 2\pi]$ . That is

$$\begin{aligned} a_0 c_0 &= 0 \\ d_k a_k - c_k b_k &= 0 \\ c_k a_k + d_k b_k &= 0. \end{aligned} \quad (46)$$

Since  $a_0, a_k, b_k \neq 0$  then  $c_0 = c_k = d_k = 0$  for all  $k = 1, \dots, N$ , as a result  $w = 0$ .

Applying Theorem 1.2 we get the uniqueness of  $F$ . ■

## References

- [1] D. L. Berman, *On the impossibility of constructing a linear polynomial operator furnishing an approximation of the order of the best approximation*, Dokl. Akad. Nauk SSSR 120 (1958), pp. 1175–1177. MR0098941 (20 #5387)
- [2] E. W. Cheney, C. R. Hobby, P. D. Morris, F. Schurer, and D. E. Wulbert, *On the minimal property of the Fourier projection*, Trans. Amer. Math. Soc. 143 (1969), pp. 249–258. MR0256044 (41 #704)
- [3] E. W. Hardy, *Note on Lebesgue's constants in the theory of Fourier series*, J. London Math. Soc. 17 (1942), pp. 4–13. MR0006754 (4,36f)
- [4] Pol V. Lambert, *On the minimum norm property of the Fourier projection in  $L^1$ -spaces*, Bull. Soc. Math. Belg. 21 (1969), pp. 370–391. MR0273293 (42 #8173)

- [5] G. Lewicki and L. Skrzypek, *Chalmers-Metcalf operator and uniqueness of minimal projections*, J. Approx. Theory 148 (2007), pp. 71–91. MR2356576
- [6] S. M. Lozinski, *On a class of linear operations*, Doklady Akad. Nauk SSSR (N. S.) 61 (1948), pp. 193–196. MR0026699 (10,188c)
- [7] W. Odyniec and G. Lewicki, 1449, *Minimal projections in Banach spaces*. Springer-Verlag, Berlin, 1990, Problems of existence and uniqueness and their application. MR1079547 (92a:41021)
- [8] W. Ruess and C. Stegall, *Extreme Points in Duals of Operator Spaces*, Math. Ann. 261 (1982), pp. 535–546. MR682665 (84e:46007)
- [9] B. Shekhtman and L. Skrzypek, *Norming points and unique minimality of orthogonal projections*, Abstr. Appl. Anal. (2006), pp. Art. ID 42305, 17. MR2211664 (2006k:46043)
- [10] B. Shekhtman, L. Skrzypek, Uniqueness of minimal projections onto two-dimensional subspaces, Studia Math. 168 (3) (2005) 273–284.
- [11] B. Shekhtman and L. Skrzypek, *On the non-uniqueness of minimal projection in  $L_p$  spaces*, J. Approx. Theory, doi:10.1016/j.jat.2008.08.006

## Object Registration Using Graph Representations of Images

*Tamir Nave, Joseph M. Francos and Rami Hagege*

Electrical and Computer Engineering Department  
Ben-Gurion University  
Beer Sheva 84105, Israel

### ABSTRACT

We consider the problem of object registration based on a set of known template images. The proposed solution employs a weighted graph representation of images, and a method that reduces the high dimensional problem of evaluating the orbit created by applying the set of all transformations in the group to a template, into a set of linear equations. The method yields a very large number of independent linear constraints that enable an explicit parametric estimation. Mathematical proof of the method is presented as well as analyzes and experiments that demonstrate its robustness.

**Index Terms**— Image registration, Image recognition, Parameter estimation, Nonlinear estimation, Graph representation of images, Multidimensional signal processing

### 1. INTRODUCTION

This paper is concerned with the general problem of automatic image registration based on a set of known templates. More specifically the paper presents the notion of compactly storing the topology of the image in a graph representation, and based on this representation, proposes an algorithmic solution to the registration problem. The fundamental setting of the problem and common approaches are provided in [1]-[4]. There are two key elements in a deformable template representation: A typical element (the template); and a family of transformations and deformations which when applied to the typical element produces other elements. The family of deformations considered in this paper is extremely wide: we consider differentiable homeomorphisms having a continuous and differentiable inverse, where the derivative of the inverse is also continuous.

Thus each template is associated with its orbit, induced by the group action on the template. Hence, given measurements of an observed object (for example, in the form of an image) registration becomes the procedure of finding the group element that minimizes some metric with respect to the observation. Theoretically, in the absence of noise, the solution to the registration problem is obtained by applying each of the deformations in the group to the template, followed by comparing the result to the observed realization. However, as the

number of such possible deformations is infinite, this direct approach is computationally prohibitive. Hence, more sophisticated methods are essential. The analysis and the algorithmic solution derived in this paper enable a rigorous treatment of the homeomorphism estimation problem in a wide range of applications.

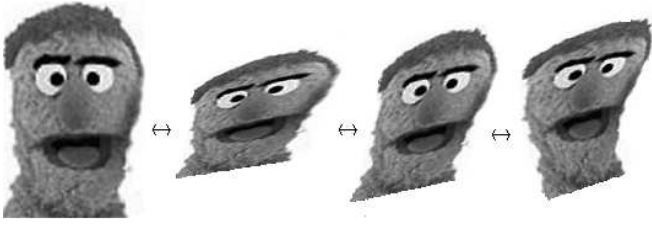
The center of the proposed solution is a method that reduces the high dimensional problem of evaluating the orbit created by applying the set of all possible homeomorphic transformations in the group to the template, into a problem of analyzing a function in a low dimensional Euclidian space. In general, an explicit modeling of the homeomorphisms group is impossible. We therefore choose to solve this problem by focusing on subsets of the homeomorphisms group which are also subsets of vector spaces. This may be regarded as an approximation the homeomorphism using polynomials, based on the denseness of the polynomials in the space of continuous functions with compact support. In this setting, the problem of estimating the parametric model of the deformation is solved by a *linear* system of equations in the low dimensional Euclidian space.

More specifically, consider the problem given by  $h(x_1, x_2) = f(\phi(x_1, x_2))$  where  $\phi(x_1, x_2) = (\phi_1(x_1, x_2), \phi_2(x_1, x_2))$ . In the problem setting considered here  $h$  and  $f$  are given while  $\phi$  should be estimated. In [5],[6],[7] we analyzed this problem and derived estimation algorithms of the deformation  $\phi$ , using non linear functionals employed to construct linear constraints on the parameters of the homeomorphisms. This paper presents a discrete graph representation of images and proposes a generalization of these functionals. The new functionals employ the graph representation of the image. The concept of graph representation of images stems from digital topology [9], and has been employed in a verity of applications [10], [11], [13], [14]. The common concept behind these representations is to have the topology of the image compactly packed in a graph structure.

### 2. GRAPH REPRESENTATION OF AN IMAGE

Observe a continuous 2D image of an object and allow it to deform elastically into any shape as long as the deformation is a homeomorphism, (i.e. a continuous and invertible trans-

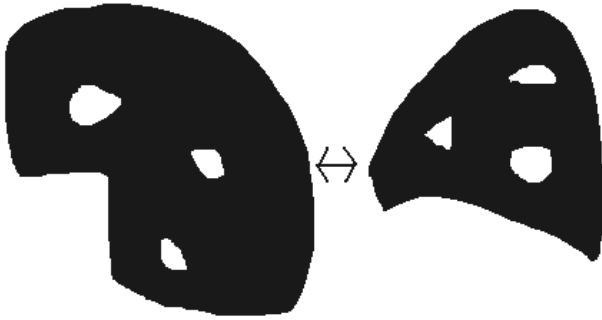
formation such that its inverse is also continuous). See Figure 1.



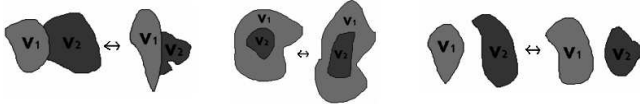
**Fig. 1.** Homeomorphisms on an image.

Obviously, the key to the solution of the any registration problem is the understanding of what are the basic invariant properties of the deformed object. Clearly, its size and contours are not; the edges, and the straight lines in it also vary due to the elastic deformation. The set of colors of the deformed image may also change due to varying illumination conditions. Yet, there is a fundamental property that remains invariant and this is the object topology.

It is known that homeomorphisms preserve connectivity and the number of holes in a set. (e.g., Figures 2 , 3)

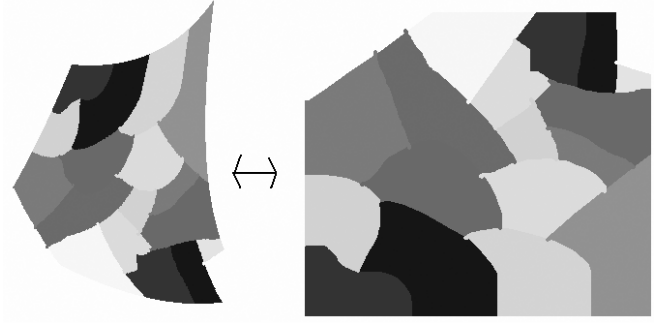


**Fig. 2.** Homeomorphisms preserve the number of holes in a set, not their shape.



**Fig. 3.** The topology of any two sets is maintained, after homeomorphisms.

Now, let us treat an image as a collection of colored patches. Each patch is a connected set in  $R^2$  adjacent to different patches. Hence, the neighboring patches of any patch remain its neighbors following any homeomorphism. (See, e.g., Figure 4)



**Fig. 4.** The topology of a complex structure of sets is invariant to homeomorphisms.

Denoting the set of all bounded and measurable functions with compact support from  $X \subset R^2$  into  $Y \subset R$  by  $M$ , the previous understanding leads us to define the following transformation from  $M$  into the set of weighted graphs:  $f(x, y) \rightarrow \langle E, V \rangle$  by the following rule:

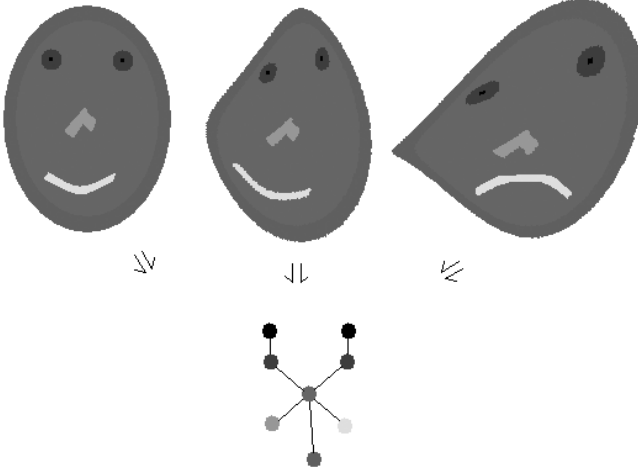
$$\begin{aligned} v \in V &\iff v \subseteq f^{-1}(i) \\ \langle v_1, v_2 \rangle \in E &\iff cl(v_1) \cap cl(v_2) \neq \emptyset \end{aligned} \quad (1)$$

where  $i \in Im\{f\}$  is the weight of the vertex  $v$ ,  $v \subseteq X$  is a connected set, and  $cl$  means the closure of a set. The representation is non bijective: An image is represented by one graph only, however, a single graph represents many images.

The motivation to represent an image by a weighted graph is that graph representation is homeomorphism-invariant. (See, e.g. Figure 4, 5). The main disadvantage of this representation is that apart from geometric deformation, real images suffer from intensity variations that may change their graph representation.

### 3. PROBLEM DESCRIPTION

In this section we shall briefly set the mathematical framework we adopt in order to formalize the analysis of the deformation estimation problem. This framework enables accurate representation and analysis of our problem, leading to rigorous criteria on the existence and uniqueness of the solution, and under some mild restrictions to be explained below to the derivation of an explicit solution. We note that due to the inherent physical properties of the problem, it is natural to model and solve it in the continuous domain. Inherently, the mapping  $\phi$  of  $R^2$  into itself is of a continuous nature, as is the physical phenomenon of geometric deformation of real-life objects it represents. Thus, if we impose a discrete model (e.g.,  $(x_1, x_2) \in Z^2$ ), we find that, in general, the natural  $\phi$  to consider is incompatible (as for “almost all”  $(x_1, x_2) \in Z^2$ ,  $(\phi_1(x_1, x_2), \phi_2(x_1, x_2)) \notin Z^2$ ). Thus, the problem and its solution are formulated in the continuous domain, while the sampling and quantization effects that accompany the digital



**Fig. 5.** Several instances of the same object that are distinct from each other by homeomorphisms, however they are all represented by the same graph.

implementation of the method, are handled as noise contributions.

### 3.1. Group Theory Setting

Let  $M$  denote the space of compact support, bounded, and Lebesgue measurable (or more simply, integrable) functions from  $R^2$  to  $R$ . Let  $\mathbf{x}$  be some vector in  $R^2$ .

Let  $G$  be a group representing the set of deformations the objects may undergo.  $G$  is said to act as a transformation group on  $M$  if there is a mapping  $G \times M \rightarrow M$ , denoted by  $(\phi, f) \mapsto f \circ \phi = f(\phi(\mathbf{x}))$  such that  $(f \circ \phi_1) \circ \phi_2 = f \circ (\phi_1 \circ \phi_2)$  for every  $\phi_1, \phi_2 \in G$  and  $f \in M$ ; and if  $f \circ e = f$  for all  $f \in M$ , where  $e$  is the identity element of  $G$ .

For a given  $f \in M$ , the set  $\{f \circ \phi : \phi \in G\}$  is called the orbit of  $f$ . It is the entire set of possible observations on the object – the result of applying to it any of the deformations in the group.

The stabilizer of the function  $f \in M$  with respect to the group  $G$  is the set of group elements  $\phi \in G$  such that  $f \circ \phi = f$ , i.e., the set of group elements that map  $f$  to itself.

Thus the group  $G$  naturally defines an equivalence relation on  $M$  in terms of the orbits of  $M$  induced by the action of  $G$ : Any two functions  $h$  and  $f$  are equivalent if they are on the same orbit, i.e., if there exists some  $\phi \in G$  such that  $f \circ \phi = h$ .

Let  $M_G \subseteq M$  be the subset of functions in  $M$  with no group symmetry, i.e., the set of functions in  $M$  whose stabilizer is trivial and includes only  $e$ , the identity element of  $G$ . Thus,  $M_G$  is the subset of functions in  $M$  where *uniqueness of the solution* to the defined problem is guaranteed in the sense that if  $h, f \in M_G$  such that they are on the same orbit, then there exists a *single*  $\phi$  such that  $f \circ \phi = h$ . In [8] we show that  $M_G$  is dense in  $M$  in the  $L^2$  norm. In contrast, examples

of functions that don't belong to  $M_G$ , where  $G$  is the affine group, include any constant function defined on all of  $R^2$ ; any periodic function defined on all of  $R^2$ ; and functions with radial symmetry, such as a circle (as  $SO_2(R) \subset GL_2(R)$ ). Note however that functions with compact support are not translation nor scale invariant.

### 3.2. Problem Statement

In the following we assume that  $G$  is the group of differentiable homeomorphisms such that each element of  $G$  has a continuous and differentiable inverse, where the derivative of the inverse is also continuous. The group  $G$  lies in the norm space  $C(X)$  of continuous real-valued functions of  $X$ . By the above assumption every  $\phi^{-1}, (\phi^{-1})' \in C(X)$ . Since  $C(X)$  is a normed separable space, there exists a countable set of basis functions  $\{e_i\} \subset C(X)$ , such that for every  $\phi \in G$ ,

$$(\phi^{-1})'(x) = \sum_i b_i e_i(x). \quad (2)$$

In other words, it is assumed that every element in the group and its derivative can be represented as a convergent series of basis functions of the separable space  $C(X)$ . Our goal then, is to obtain the expansion of  $\phi^{-1}(x)$  with respect to the basis functions  $\{e_i(x)\}$ . In practice, the series (2) is replaced by a finite sum, i.e., we have  $1 \leq i \leq m$ .

Given two bounded, Lebesgue measurable functions  $h, g \in M_G$  with compact supports, such that

$$h(\mathbf{x}) = f(\phi(\mathbf{x})), \quad \phi \in G, \quad \mathbf{x} \in R^2 \quad (3)$$

the problem is to find the deformation  $\phi$ . As indicated above, the direct approach for solving the problem of finding the parameters of the unknown transformation  $\phi \in G$  is to apply the set of all possible transformations, (i.e., every element of  $G$ ), to the given template  $f$ , thus evaluating the entire orbit of  $f$ . Since  $h$  and  $f$  are homeomorphic, one of the points on the orbit represents the action of the desired group element  $\phi$ . Nevertheless, since  $\phi$  is modeled by  $m$  parameters, it is clear that implementation of such a search on the orbit requires a search over an  $m$ -dimensional manifold embedded in an infinite dimensional function space, which is infeasible.

In this paper we show that the problem of finding the parameters of the unknown elastic transformation, whose direct solution requires a highly complex search in a function space, can be formulated as an *explicit parameter estimation problem*. Moreover, it is shown that the original problem can be formulated in terms of an *equivalent* problem which is expressed in the form of a *linear* system of equations. From every subgraph of the graph that represents the template we obtain a linear constraint in the unknown parameters of the transformation. A solution of this linear system of equations provides the unknown transformation parameters. In Section 4 we show how the problem of finding the parametric model

of the deformation can be transformed using a set on non-linear graph-based functionals into a set of *linear* equations which is then solved for the transformation parameters. However, before getting in Section 4 into the details of this new representation of the problem, we shall briefly elaborate on the mathematical construction that enables it.

### 3.3. The Mathematical Structure and the Fundamental Commutative Property

Recall that  $M_G$  is the space of compact support, bounded, measurable functions, with no group symmetry. Let  $L_\phi$  be the mapping from  $M_G$  to itself induced by the group  $G$ , such that  $L_\phi(f(x)) = f(\phi(x))$  for every  $f \in M_G$  and every  $\phi \in G$ . Since  $L_\phi(af_1(x) + bf_2(x)) = aL_\phi(f_1(x)) + bL_\phi(f_2(x))$ , we have that  $L_\phi$  is a linear operator. Thus, the problem we address can be restated as follows: Given the pair  $L_\phi(f(x)), f(x)$  find the linear operator  $L_\phi$ .

Towards this goal, let us define an operator  $w$  such that:

$$w : M_G \times G^* \rightarrow M_G \quad (4)$$

where  $G^*$  is the set of weighted graphs in which the weight of each vertex consists of two elements of  $Y$ :

$$G^* = \{ \langle V, E, \psi \rangle, \psi : V \rightarrow Y^2; \psi = (\psi^1, \psi^2) \} \quad (5)$$

The operator  $w$  acts on an image  $f \in M_G$  and a graph  $g_p \in G^*$  by finding all of the graphs  $g_p$  as subgraphs of the graph that represents the image  $f$ . To each detected vertex  $v$  with an intensity  $\psi^1(v)$  the operator assigns a new value  $\psi^2(v)$  and eliminates the graph vertices that weren't found. Thus a new image is formed with non zeros values only at locations that satisfy the definition of the subgraph  $g_p$ . This mapping, denoted by  $M_{g_p} : M \rightarrow M$ , is defined as follows:  $M_{g_p}(f(x)) = w(f(x), g_p)$ ,  $g_p \in G^*$ . The fundamental property being exploited in this paper in order to reduce the original high dimensional problem to an equivalent problem that is linear in the unknown transformation parameters is the commutative property of the left composition operator  $w$  and the right composition operator  $\phi$ , stated explicitly in the next theorem:

**Theorem 1.** *Let  $f, h \in M_G$  and  $g_p \in G^*$ . Then  $L_\phi(f(x)) = h(x)$  implies that  $L_\phi(w(f(x), g_p)) = w(h(x), g_p)$ . In a more concise form:  $M_{g_p}(L_\phi) = L_\phi(M_{g_p})$*

*Proof.*  $[M_{g_p}(L_\phi)](f(x)) = M_{g_p}(f(\phi(x))) = w(f(\phi(x)), g_p) = L_\phi(w(f(x), g_p)) = [L_\phi(M_{g_p})](f(x))$   $\square$

Thus, knowing how  $L_\phi$  acts on some function  $f$ , we know the action of  $L_\phi$  on any function  $w(f(x), g_p)$ , for any  $g_p \in G^*$ .

## 4. LINEAR CONSTRAINTS FROM DEFORMED IMAGE

To simplify the notation and the accompanying discussion we present the solution for the case where the observed signals are one-dimensional. The derivation for higher dimensions follows along similar lines. Consider the problem formulated in (3) and let  $z = \phi(x)$ . Then  $\phi^{-1}(z) = x$ , and hence

$$(\phi^{-1})'(z)dz = dx \quad (6)$$

Let us choose any  $p$  elements  $\{g_p\}_{p=1}^P \subseteq G^*$ , and as we show next, these elements are employed to translate the identity relation (3) into a set of  $P$  equations:

$$\begin{aligned} \int_{-\infty}^{\infty} w(h(x), g_p)dx &= \int_{-\infty}^{\infty} w(f(\phi(x)), g_p)dx \\ &= \int_{-\infty}^{\infty} (\phi^{-1}(z))' w(f(z), g_p)dz \\ &= \sum_{i=1}^m b_i \int_{-\infty}^{\infty} e_i(x) w(f(x), g_p)dx \\ &\quad p = 1, \dots, P \end{aligned} \quad (7)$$

Rewriting (7) in a matrix form we have

$$\begin{bmatrix} \int w(h, g_1) \\ \vdots \\ \int w(h, g_p) \end{bmatrix} = \begin{bmatrix} \int e_1 w(f, g_1) & \dots & \int e_m w(f, g_1) \\ \vdots & \ddots & \vdots \\ \int e_1 w(f, g_p) & \dots & \int e_m w(f, g_p) \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} \quad (8)$$

Based on the fact that the operator  $w$  is homeomorphism invariant, we have the following theorem:

**Theorem:** The homeomorphism  $\phi$  satisfying the parametric model defined in (2) is uniquely determined iff the matrix

$$\begin{bmatrix} \int e_1 w(f, g_1) & \dots & \int e_m w(f, g_1) \\ \vdots & \ddots & \vdots \\ \int e_1 w(f, g_p) & \dots & \int e_m w(f, g_p) \end{bmatrix} \quad (9)$$

is full rank.

Thus, provided that  $\{g_p\}_{p=1}^P$  are chosen such that (9) is full rank, the system (8) (in the absence of noise we take  $P = m$ ) can be solved for the parameter vector  $[b_1, \dots, b_m]$ . It is clear that in the absence of noise, any set of weighted graphs  $\{g_p\}_{p=1}^m$  such that (9) is full rank is equally optimal.

## 5. NUMERICAL EXAMPLES

The following example illustrates the proposed solution for elastic image registration. The template was taken to be an

RGB image of dimensions  $314 \times 314$ , and the observation is an elastic deformed version of it (Figure (6)) where the deformation function is:

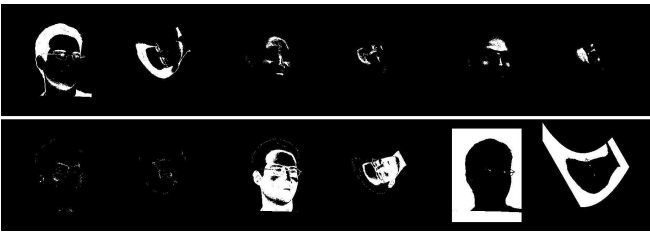
$$\phi(x, y) = (-0.75x - 1.29y - 1.35x^2, 1.29x - 0.75y)$$



**Fig. 6.** Template and elastically deformed observation.

The estimation does not involve any search scheme, but only the application of the graph representation of each image and our linear model. Figure 7 depicts the operation of functionals that are based on all sub-graphs with one vertex only. In this example each vertex corresponds to a subset of the image color space that consist of  $100^3$  colors out of the entire RGB color space of  $256^3$  colors. The label assigned to each vertex is an indicator function of the corresponding color cube. Figure 8 depicts the operation of functionals that are based on all sub-graphs with two vertices, using indicator functions as vertices' labels. Figure 9 depicts the operation of functionals that are based on another type of one-vertex sub-graph where different Gaussian functions of the observed intensities are employed as vertices' labels. The estimated deformation obtained using the constraints derived using all these subgraphs is given by:

$$\phi(x, y) = (-0.73x - 1.25y - 1.3x^2, 1.29x - 0.76y - 0.01x^2).$$



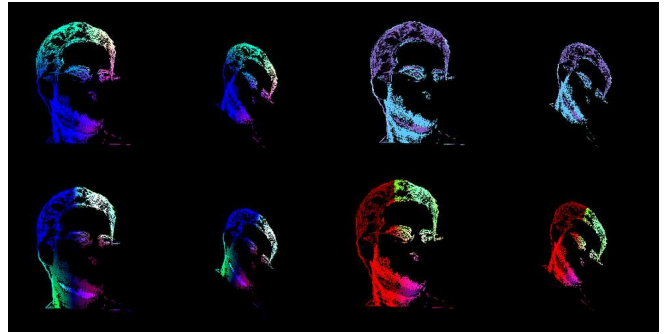
**Fig. 7.** The operator application to the template and observation using all sub-graphs with one vertex.

## 6. CONCLUSIONS

In this paper we have considered the problem of finding the transformation relating a given observation on a planar ob-



**Fig. 8.** The operator application on the template and observation based on all sub-graphs with two vertices.



**Fig. 9.** The operator application on the template and observation based on one vertex sub-graph with Gaussian weighted labels on its vertices.

ject with some pre-chosen template of this object. The direct approach for estimating the transformation is to apply each of the deformations in the group to the template in a search for the deformed template that matches the observation. The notion that a weighted graph represents the topology of images was presented. A method that employs a set of non-linear graph-based functionals to replace the original high dimensional problem by an *equivalent linear problem*, expressed in terms of the unknown transformation parameters, was derived. The resulting method is explicit and global. It deals with any elastic deformation, and the obtained map  $\varphi = H(h, f)$  is continuous and involves only elementary linear analysis in the same dimension as that of the group model.



## 7. REFERENCES

- [1] U. Grenander, *General Pattern Theory*, Oxford University Press, 1993.
- [2] L. Brown, "A Survey of Image Registration Techniques," *ACM Comput. Surv.*, vol. 24, no. 4, pp. 326-376, 1992.
- [3] B. Zitova and J. Flusser, "Image Registration Methods: A Survey," *Image, Vis. Comp.*, vol. 21, pp. 977-1000, 2003.
- [4] M. Miller and L. Younes, "Group Actions, Homeomorphisms, and Matching: A General Framework," *Int. Jou. Comp. Vision*, vol. 41, pp. 61-84, 2002.
- [5] R. Hagege and J. M. Francos, "Parametric Estimation of Two-Dimensional Affine Transformations," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Montreal 2004.
- [6] J. M. Francos, R. Hagege and B. Friedlander, "Estimation of Multi-Dimensional Homeomorphisms for Object Recognition in Noisy Environments," *Proc. Thirty Seventh Asilomar Conference on Signals, Systems, and Computers*, 2003.
- [7] R. Hagege and J. M. Francos, "Linear Estimation of Sequences of Multi-Dimensional Affine Transformations," *Int. Conf. Acoust., Speech, Signal Processing*, Toulouse, 2006.
- [8] R. Hagege and J. M. Francos, "Parametric Estimation of Affine Transformations: An Exact Linear Solution", submitted for publication.
- [9] U. Eckhardt and L. Latecki, "Digital Topology", Research Trends, Council of Scientific Information, Vilayil Gardens, Trivandrum, India, 1994.
- [10] P. L. Bazin, L. M. Ellingsen, and D. L. Pham, "Digital Homeomorphisms in Deformable Registration", *Information Processing in Medical Imaging* p.211-222 Volume 4584/2007 Springer Berlin / Heidelberg.
- [11] L. G. Nonato, A. M. da Silva Junior, J. Batista, and O. M. Bruno, "Circulation and Topological Control in Image Segmentation", *Progress in Pattern Recognition, Image Analysis and Applications* p.377-391 Volume 3773/2005 Springer Berlin / Heidelberg.
- [12] E. Decencire and M. Bilodeau, "Downsampling of Binary Images Using Adaptive Crossing Numbers", "40 Years On: Mathematical Morphology", Volume 30 p.279-288 Springer Netherlands.
- [13] A. Charnoz, V. Agnus, and L. Soler, "Portal Vein Registration for the Follow-Up of", *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2004* Volume 3216/2004 p. 878-886 Springer Berlin / Heidelberg.
- [14] S. Todorovic and N. Ahuja, "Region-Based Hierarchical Image Matching", *International Journal of Computer Vision*, Volume 78, NO. 1, June, 2008 p. 47-66.

# Trajectory Tubes of Nonlinear Differential Inclusions and State Estimation Problems

Tatiana F. Filippova

Department of Optimal Control  
Institute of Mathematics and Mechanics  
Russian Academy of Sciences  
16 S. Kovalevskaya Str., Ekaterinburg 620219, Russia  
*e-mail: [ftf@imm.uran.ru](mailto:ftf@imm.uran.ru)*

## Abstract

The paper is devoted to state estimation problems for nonlinear dynamic systems with system states being compact sets. The studies are motivated by the theory of dynamical systems with unknown but bounded uncertainty without its statistical description. The trajectory tubes of differential inclusions are introduced as the set-valued analogies of the classical isolated trajectories of uncertain dynamical systems. Applying results related to discrete-time versions of the funnel equations and techniques of ellipsoidal estimation theory developed for linear control systems we present new approaches that allow to find the outer and inner estimates for such set-valued states of the uncertain nonlinear control system. Numerical simulations are also given.

**Key word:** Differential inclusions; Uncertain dynamic system; State constraints; Viability theory; Trajectory tube; Funnel equations; State estimation; Ellipsoidal approach.

## 1 Introduction

The topics of this paper come from the theory of dynamical control systems with unknown, but bounded uncertainties (the case of the so-called "set-membership" description of uncertainties) [3, 5, 11, 14, 15, 17, 18]. The motivations for these studies come from applied areas ranged from engineering problems in physics to economics as well as to ecological modelling. The paper presents recent results in the theory of tubes of solutions (trajectory tubes) to differential control systems modelled by nonlinear differential inclusions with uncertain parameters or functions.

We will start by introducing the following basic notations. Let  $R^n$  be the  $n$ -dimensional Euclidean space and  $x'y$  be the usual inner product of  $x, y \in R^n$  with the prime as a transpose, with  $\|x\| = (x'x)^{1/2}$ . Denote  $\text{comp } R^n$  to be the variety of all compact subsets  $A \subseteq R^n$  and  $\text{conv } R^n$  to be the variety of all compact convex subsets  $A \subseteq R^n$ . Let us denote the variety of all closed convex subsets  $A \subseteq R^n$  by the symbol  $\text{clconv } R^n$ .

Consider the ordinary differential equation

$$\dot{x} = f(t, x, u) \quad (1)$$

with function  $f : T \times R^n \times R^n \rightarrow R^n$  measurable in  $t$  and continuous in the other variables. Here  $x$  stands for the state space vector,  $t$  stands for time ( $t \in T = [t_0, t_1]$ ) and  $u$  is a control or a disturbance. The variables  $u$  in (1) are assumed to be bounded

$$u \in Q(t, x) \quad (2)$$

where  $Q(t, x)$  is a set-valued map ( $Q : T \times R^n \rightarrow \text{comp } R^n$ ) measurable in  $t$  and continuous in  $x$ . The given data allows to consider a set-valued function

$$F(t, x) = \bigcup \{ f(t, x, u) \mid u \in Q(t, x) \} \quad (3)$$

and further on, a differential inclusion [2, 4, 8]

$$\dot{x} \in F(t, x) \quad (4)$$

that reflects the variety of all models of type (1)-(2).

Let us assume that the initial condition to the system (1) (or to the differential inclusion (4)) is unknown also but bounded

$$x(t_0) = x^0, \quad x^0 \in X^0 \in \text{comp } R^n \quad (5)$$

One of the principal points of interest of the theory of control under uncertainty conditions [14] is to study the set of all solutions  $x[t] = x(t, t_0, x^0)$  to (1)-(5) (respectively, (4)-(5)) and furthermore the subset of those trajectories  $x[t] = x(t, t_0, x^0)$  that satisfy both (4)-(5) and a restriction on the state vector (the "viability" constraint [1])

$$x[s] \in Y(s), \quad s \in [t_0, t] \quad (6)$$

where  $Y(\cdot)$  ( $Y(t) \in \text{conv } R^p$ ) is a convex compact valued multifunction.

The viability constraint (6) may be induced by state constraints defined for a given plant model or by the so-called measurement equation [14]

$$y(t) = G(t)x + w, \quad (7)$$

where  $y$  is the measurement,  $G(t)$  — a matrix function,  $w$  — the unknown but bounded "noise" and

$$w \in Q(t), \quad Q(t) \in \text{comp } R^p.$$

The problem consists in describing the set  $X[\cdot] = \{x[\cdot] = x(\cdot, t_0, x^0)\}$  of solutions to the system (4)- (5) (or to the system (4)- (6), that is the viable solution bundle or "viability bundle"). The point of special interest is to describe the  $t$  – cross-section  $X[t]$  of this set that is actually the attainability domain of system (1), (4), (5) at the moment  $t$ . The set  $X[t]$  may be considered also as the set-valued estimate of the unknown state  $x(t)$  of the system of relations (4), (5) and (6). This estimate  $X[t]$  as a set-valued function of  $t \in [t_0, t_1]$  is called *the viability tube* or *the viable solution tube*. The viability tubes were considered in some aspects in the theory of differential games [12, 13] and in nonlinear control synthesis problems [16].

The paper deals with the problems of control and state estimation for a dynamical control system described by differential inclusions with unknown but bounded initial state. The solution to the differential system is studied through the techniques of trajectory tubes with their cross-sections  $X(t)$  being the reachable sets at instant  $t$  to control system.

Basing on the well-known results of ellipsoidal calculus developed for linear uncertain systems we present the modified state estimation approaches which use the special nonlinear structure of the control system and simplify calculations. Examples and numerical results related to procedures of set-valued approximations of trajectory tubes and reachable sets are also presented.

## 2 Preliminaries

We assume that the notions of continuity and measurability of set-valued maps are taken in the sense of [4].

Consider the differential inclusion (4), where  $x \in R^n$ ,  $F$  is a continuous multivalued map ( $F : [t_0, t_1] \times R^n \rightarrow \text{conv}R^n$ ) that satisfies the Lipschitz condition with constant  $L > 0$ , namely

$$h(F(t, x), F(t, y)) \leq L \|x - y\|, \quad \forall x, y \in R^n$$

where  $h(A, B)$  is the Hausdorff distance for  $A, B \subseteq R^n$ , i.e.

$$h(A, B) = \max \{h^+(A, B), h^-(A, B)\},$$

with  $h^+(A, B), h^-(A, B)$  being the Hausdorff semidistances between the sets  $A, B$ ,

$$h^+(A, B) = \sup\{d(x, B) \mid x \in A\}, \quad h^-(A, B) = h^+(B, A),$$

$$d(x, A) = \inf \{\|x - y\| \mid y \in A\}.$$

Assuming a set  $X_0 \in \text{comp}R^n$  to be given, denote  $x[t] = x(t, t_0, x_0)$  ( $t \in T = [t_0, t_1]$ ) to be a solution to (4) (an isolated trajectory) that starts at point  $x[t_0] = x_0 \in X_0$ .

We take here the Caratheodory-type trajectory  $x[\cdot]$ , i.e. as an absolutely continuous function  $x[t]$  ( $t \in T$ ) that satisfies the inclusion

$$\frac{d}{dt} x[t] = \dot{x}[t] \in F(t, x[t]) \quad (8)$$

for almost every  $t \in T$ .

We require all the solutions  $\{x[t] = x(t, t_0, x_0) \mid x_0 \in X_0\}$  to be extendable up to the instant  $t_1$  that is possible under some additional assumptions [8].

Let  $Y(t)$  be a continuous set-valued map ( $Y : T \rightarrow \text{conv}R^n$ ),  $X_0 \subseteq Y(t_0)$ .

**Definition 1** [1, 15] *A trajectory  $x[t] = x(t, t_0, x_0)$  ( $x_0 \in X_0$ ,  $t \in T$ ) of the differential inclusion (8) is called viable on  $[t_0, \tau]$  if*

$$x[t] \in Y(t) \quad \text{for all } t \in [t_0, \tau]. \quad (9)$$

We will assume that there exists at least one solution  $x^*[t] = x^*(t, t_0, x_0^*)$  of (8) (together with a starting point  $x^*[t_0] = x_0^* \in X_0$ ) that satisfies condition (9) with  $\tau = t_1$ .

Let  $\mathcal{X}(\cdot, t_0, X_0)$  be the set of all solutions to the inclusion (8) that emerge from  $X_0$  (the "trajectory bundle"). Denote  $\mathcal{X}[t] = \mathcal{X}(t, t_0, X_0)$  to be its cross-section at instant  $t$ .

The subset of  $\mathcal{X}(\cdot, t_0, X_0)$  that consists of all solutions to (8) viable on  $[t_0, \tau]$  will be further denoted as  $X(\cdot, \tau, t_0, X_0)$  (the "viable trajectory bundle") with its  $s$ -crosssections as  $X(s, \tau, t_0, X_0)$ ,  $s \in [t_0, \tau]$ . We introduce symbol  $X[\tau]$  for these crosssections at instant  $\tau$ , namely

$$X[\tau] = X(\tau, t_0, X_0) = X(\tau, \tau, t_0, X_0)$$

It is known that both maps  $\mathcal{X}(t, t_0, X_0)$ ,  $X(t, t_0, X_0)$ ,

$$\mathcal{X} : T \times T \times \text{comp}R^n \rightarrow \text{comp}R^n,$$

$$X : T \times T \times \text{comp}R^n \rightarrow \text{comp}R^n,$$

satisfy the semigroup property:

$$\mathcal{X}(t, \tau, \mathcal{X}(\tau, t_0, X_0)) = \mathcal{X}(t, t_0, X_0), \quad t_0 \leq \tau \leq t \leq t_1,$$

$$X(t, \tau, X(\tau, t_0, X_0)) = X(t, t_0, X_0), \quad t_0 \leq \tau \leq t \leq t_1,$$

and therefore define the generalized dynamic systems with set-valued trajectories [3, 20, 15]. The multivalued functions  $\mathcal{X}[t]$  and  $X[t]$  ( $t \in T$ ) will be referred to as the *trajectory tube* and *viable trajectory tube* (or *viability tube*) respectively. They may be considered as the set-valued analogies of the classical isolated trajectories constructed now under uncertainty conditions.

One of the approaches that we discuss here is related to the evolution equation of the "funnel type" that describes the dynamics of set-valued "states". The basic assumptions on set-valued map  $F(t, x)$  for the following results to be true may be found in [15, 9].

Let us consider the "equation"

$$\lim_{\sigma \rightarrow +0} \sigma^{-1} h \left( \mathcal{X}[t + \sigma], \bigcup_{x \in \mathcal{X}[t]} (x + \sigma F(t, x)) \right) = 0, \quad t \in T = [t_0, t_1] \quad (10)$$

with "initial condition"

$$\mathcal{X}[t_0] = X_0. \quad (11)$$

We can observe that this equation is the formal analogy of the ordinary differential equation when mappings  $F(t, x) = \{f(t, x)\}$  and  $\mathcal{X}[t] = \{x[t]\}$  ( $X_0 = \{x_0\}$ ) are single-valued.

**Theorem 1** [19] *The multifunction  $\mathcal{X}[t] = \mathcal{X}(t, t_0, X_0)$  is the unique set-valued solution to the evolution equation (10)-(11).*

Other versions of funnel equation (10) may be considered by substituting the Hausdorff distance  $h$  for a semidistance  $h^+$  [16]. The solution to the  $h^+$ -versions of the evolution equation may be not unique and the "maximal" one (with respect to inclusion) is studied.

Now let us consider the analogy of the funnel equation (10)-(11) but now for the viable trajectory tubes  $X[t] = X(t, t_0, X_0)$ :

$$\lim_{\sigma \rightarrow +0} \sigma^{-1} h(X[t + \sigma], \bigcup_{x \in X[t]} (x + \sigma F(t, x)) \cap Y(t + \sigma)) = 0, \quad t \in T, \quad (12)$$

$$X[t_0] = X_0. \quad (13)$$

The following result was proved in [15, 9] under assumptions of different type concerning mappings  $F(t, x)$  and  $Y(t)$ .

**Theorem 2** [9, 15, 16] *The multivalued function  $X[t] = X(t, t_0, X_0)$  is the unique solution to the evolution equation (12)-(13).*

### 3 Problem Statement

It should be noted that the exact description of reachable sets of a control system is a difficult problem even in the case of linear dynamics. The estimation theory and related algorithms basing on ideas of construction outer and inner set-valued estimates of reachable sets have been developed in [17, 5] for linear control systems.

In this paper the modified state estimation approaches which use the special quadratic structure of nonlinearity of studied control system and use also the advantages of ellipsoidal calculus [17, 5] are presented. We develop here new ellipsoidal techniques related to constructing external and internal set-valued estimates of reachable sets and trajectory tubes of the nonlinear system. Some estimation algorithms basing on combination of discrete-time versions of evolution funnel equations and ellipsoidal calculus [17, 5] are given. Examples and numerical results related to procedures of set-valued approximations of trajectory tubes and reachable sets are also presented. The applications of the problems studied in this paper are in guaranteed state estimation for nonlinear systems with unknown but bounded errors and in nonlinear control theory.

The paper deals with the problems of control and state estimation for a dynamical control system

$$\dot{x}(t) = A(t)x(t) + f(x(t)) + G(t)u(t), \quad (14)$$

$$x \in R^n, \quad t_0 \leq t \leq T,$$

with unknown but bounded initial condition

$$x(t_0) = x_0, \quad x_0 \in X_0, \quad X_0 \subset R^n, \quad (15)$$

$$u(t) \in U, \quad U \subset R^m, \quad \text{for a.e. } t \in [t_0, T]. \quad (16)$$

Here matrices  $A(t)$  and  $G(t)$  (of dimensions  $n \times n$  and  $n \times m$ , respectively) are assumed to be continuous on  $t \in [t_0, T]$ ,  $X_0$  and  $U$  are compact and convex. The nonlinear  $n$ -vector function  $f(x)$  in (14) is assumed to be of quadratic type

$$f(x) = (f_1(x), \dots, f_n(x)),$$

$$f_i(x) = x' B_i x, \quad i = 1, \dots, n, \quad (17)$$

where  $B_i$  is a constant  $n \times n$  - matrix ( $i = 1, \dots, n$ ).

Consider the following differential inclusion [8] related to (14)–(16)

$$\dot{x}(t) \in A(t)x(t) + f(x(t)) + P(t), \quad \text{for a.e. } t \in [t_0, T], \quad (18)$$

$$x(t_0) = x_0 \in X_0,$$

where  $P(t) = G(t)U$ .

Let absolutely continuous function  $x(t) = x(t, t_0, x_0)$  be a solution to (18) with initial state  $x_0$  satisfying (15). The differential system (14)–(16) (or equivalently, (18)) is studied here in the framework of the theory of uncertain dynamical systems (differential inclusions) through the techniques of trajectory tubes

$$X(\cdot, t_0, X_0) = \{x(\cdot) = x(\cdot, t_0, x_0) \mid x_0 \in X_0\} \quad (19)$$

of solutions to (14)–(16) with their  $t$ -cross-sections  $X(t) = X(t, t_0, X_0)$  being the reachable sets at instant  $t$  for control system (14)–(16).

The problem consists in describing the set  $X(\cdot) = \cup_{x_0 \in X_0} \{x(\cdot) = x(\cdot, t_0, x_0)\}$  of solutions to the differential inclusion (18) under constraint (6) (the viable trajectory tube). The point of special interest is to describe the  $t$  - cross-section  $X(t)$  of this map that is actually the attainability domain of this system at the instant  $t$ .

Basing on results of ellipsoidal calculus ([17, 5]) developed for linear uncertain systems we present here the modified state estimation approaches which use the special structure of nonlinearity of studied control system (14)–(17) and combine advantages of estimating tools mentioned above.

## 4 External Estimates of Reachable Sets and Trajectory Tubes

We denote as  $B(a, r)$  the ball in  $R^n$ ,  $B(a, r) = \{x \in R^n : \|x - a\| \leq r\}$ ,  $I$  is the identity  $n \times n$ -matrix. Denote by  $E(a, Q)$  the ellipsoid in  $R^n$ ,  $E(a, Q) = \{x \in R^n : (Q^{-1}(x - a), (x - a)) \leq 1\}$  with center  $a \in R^n$  and symmetric positive definite  $n \times n$ -matrix  $Q$ . For any  $n \times n$ -matrix  $Q$  denote its track as  $\text{Tr } Q$  and its determinant as  $|Q|$ .

### 4.1 Outer Ellipsoidal Bounds

The approach presented here uses the techniques of ellipsoidal calculus developed for linear control systems [5, 17]. It should be noted that external ellipsoidal approximations of trajectory tubes may be chosen in various ways and several minimization criteria are well-known. We consider here the ellipsoidal techniques related to construction of external estimates with minimal volume (details of this approach and motivations for linear control systems may be found in [5, 17]).

Assume here that  $P(t) = E(a, Q)$  in (18), matrices  $B_i$  ( $i = 1, \dots, n$ ) are symmetric and positive definite,  $A(t) \equiv A$ . We may assume that all trajectories of the system (18)-(15) belong to a bounded domain  $D = \{x \in R^n : \|x\| \leq K\}$  where the existence of such constant  $K > 0$  follows from classical theorems of the theory of differential equations and differential inclusions [8].

From the structure (17) of the function  $f$  we have two auxiliary results. Their proofs are based on the algebraic properties of quadratic forms and are omitted here.

**Lemma 1** *The following estimate is true*

$$\|f(x)\| \leq N, \quad N = K^2 \left( \sum_{i=1}^n \lambda_i^2 \right)^{1/2},$$

where  $\lambda_i$  is the maximal eigenvalue for matrix  $B_i$  ( $i = 1, \dots, n$ ).

**Lemma 2** *For all  $t \in [t_0, T]$  the inclusion*

$$X(t) \subset X^*(t)$$

*holds where  $X^*(\cdot)$  is a trajectory tube of the linear differential inclusion*

$$\dot{x} \in Ax + B(c, \sqrt{n}N/2), \quad x_0 \in X_0, \quad (20)$$

where  $c = \{N/2, \dots, N/2\} \in R^n$ .

The following theorem gives the external estimate of the trajectory tube  $X(t)$  of the differential inclusion (18).



**Theorem 3** Let  $X_0 = B(0, r)$ ,  $r \leq K$  and

$$t_* = \min \left\{ \frac{K-r}{\sqrt{2}M} ; \frac{1}{L} ; T \right\}.$$

Then for all  $t \in [t_0, t_*]$  the following inclusion is true

$$X(t, t_0, X_0) \subset E(a^+(t), Q^+(t)), \quad (21)$$

where

$$M = K\sqrt{\lambda} + N + P, \quad P = \left( \sum_{i=1}^n a_i^2 \right)^{1/2} + \sqrt{\tilde{\lambda}},$$

$$L = \sqrt{\lambda} + 2K \left( \sum_{i=1}^n \lambda_i^2 \right)^{1/2},$$

with  $\lambda$ ,  $\lambda_i$  and  $\tilde{\lambda}$  being the maximal eigenvalues of matrices  $AA'$ ,  $B_i$  ( $i = 1, \dots, n$ ) and  $Q$  respectively, and vector function  $a^+(t)$  and matrix function  $Q^+(t)$  satisfy the equations

$$\dot{a}^+ = Aa^+ + a + c, \quad a^+(t_0) = 0 \quad (22)$$

$$\begin{aligned} \dot{Q}^+ &= AQ^+ + Q^+A^T + qQ^+ + q^{-1}Q^*, \\ q &= \{n^{-1} \text{Tr}((Q^+)^{-1}Q^*)\}^{1/2}, \\ Q^+(t_0) &= Q_0 = r^2I. \end{aligned} \quad (23)$$

Here

$$Q^* = (p^{-1} + 1)\tilde{Q} + (p + 1)Q, \quad \tilde{Q} = \frac{nN^2}{2}I, \quad (24)$$

and  $p$  is the unique positive solution of the equation

$$\sum_{i=1}^n \frac{1}{p + \alpha_i} = \frac{n}{p(p+1)}, \quad (25)$$

with  $\alpha_i \geq 0$  ( $i = 1, \dots, n$ ) being the roots of the following characteristic equation

$$|\tilde{Q} - \alpha Q| = 0. \quad (26)$$

Proof. Applying Lemmas 1-2 and the ellipsoidal techniques [5, 17] and comparing the inclusions (18) and (20) we come to the relation (21).

**Example 1.** Consider the following control system

$$\begin{cases} \dot{x}_1 &= 6x_1 + u_1, \\ \dot{x}_2 &= x_1^2 + x_2^2 + u_2, \end{cases} \quad 0 \leq t \leq T, \quad (27)$$

$$X_0 = B(0, 1), \quad P(t) = B(0, 1), \quad T = 0.15, \quad K = 2.6. \quad (28)$$

Results of computer simulations based on the above theorem for this system are given at Fig. 1.

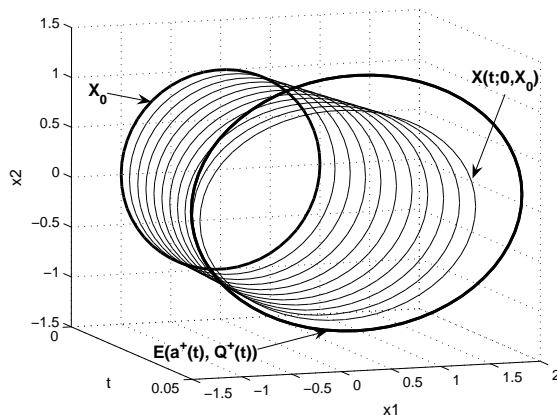


Figure 1: Reachable sets  $X(t, t_0, X_0)$  and their outer ellipsoidal estimates  $E(a^+(t), Q^+(t))$  (here  $t_* = \sqrt{2}/29.2$ ).

## 4.2 Upper Ellipsoidal Bounds via Funnel Equations

Let us discuss the estimation approach based on techniques of evolution funnel equations. Consider the following system

$$\dot{x} = Ax + \tilde{f}(x)d, \quad x_0 \in X_0, \quad t_0 \leq t \leq T, \quad (29)$$

where  $x \in R^n$ ,  $\|x\| \leq K$ ,  $d$  is a given  $n$ -vector and a scalar function  $\tilde{f}(x)$  has a form  $\tilde{f}(x) = x'Bx$  with a symmetric and positive definite matrix  $B$ .

Note that the direct application of funnel equations for finding trajectory tubes  $X(t)$  is very difficult because it takes a huge amount of computations based on grid techniques. The following theorem related to our special case of nonlinearity presents an easy computational tool to find estimates of  $X(t)$  by step-by-step procedures. For a simpler case of system nonlinearities the approach was presented in [10].

**Theorem 4** *Let  $X_0 = E(a, k^2 B^{-1})$  with  $k \neq 0$ . Then for all  $\sigma > 0$  the following inclusion holds*

$$X(t_0 + \sigma, t_0, X_0) \subseteq E(a(\sigma), Q(\sigma)) + o(\sigma)B(0, 1), \quad (30)$$

where

$$a(\sigma) = a + \sigma(Aa + a'Ba \cdot d + k^2 d), \quad (31)$$

$$Q(\sigma) = k^2(I + \sigma R)B^{-1}(I + \sigma R)', \quad R = A + 2da'B \quad (32)$$

and  $\lim_{\sigma \rightarrow +0} \sigma^{-1}o(\sigma) = 0$ .

Proof. The funnel equation for (29) is

$$\lim_{\sigma \rightarrow +0} \sigma^{-1} h(X(t + \sigma, t_0, X_0), \bigcup_{x \in X(t, t_0, X_0)} \{x + \sigma(Ax + \tilde{f}(x)d)\}) = 0, \quad t \in [t_0, T], \quad X(t_0, t_0, X_0) = X_0. \quad (33)$$

If  $x_0 \in \partial X_0$  where  $\partial X_0$  means the boundary of  $X_0$ , we have

$$\tilde{f}(x_0) = k^2 + 2a'Bx - a'Ba$$

and from (33) we have also

$$\bigcup_{x_0 \in \partial X_0} \{(I + \sigma A)x_0 + \sigma \tilde{f}(x_0)d\} = \bigcup_{x_0 \in \partial X_0} \{(I + \sigma R)x_0 + \sigma(k^2 - a'Ba)d\}. \quad (34)$$

Note that if the ellipsoid in (39) gives the tube estimate for the system with  $\partial X_0$  as starting set, then also for the system with  $X_0$  as starting set. Applying Theorem 1 and taking into account the equality (5) and the above remark we come to the estimate (39).

We may formulate now the following scheme that gives the external estimate of trajectory tube  $X(t)$  of the system (29) with given accuracy.

**Algorithm 1.** Subdivide the time segment  $[t_0, T]$  into subsegments  $[t_i, t_{i+1}]$  where  $t_i = t_0 + ih$  ( $i = 1, \dots, m$ ),  $h = (T - t_0)/m$ ,  $t_m = T$ .

- Given  $X_0 = E(a, k_0^2 B^{-1})$  with  $k_0 \neq 0$ , define  $X_1 = E(a_1, Q_1)$  from Theorem 4 for  $a_1 = a(\sigma)$ ,  $Q_1 = Q(\sigma)$ ,  $\sigma = h$ .
- Find the smallest constant  $k_1$  such that

$$E(a_1, Q_1) \subset \tilde{X}_1 = E(a_1, k_1^2 B^{-1}),$$

and it is not difficult to prove that  $k_1^2$  is the maximal eigenvalue of the matrix  $B^{1/2}Q_1B^{1/2}$ .

- Consider the system on the next subsegment  $[t_1, t_2]$  with  $E(a_1, k_1^2 B^{-1})$  as the initial ellipsoid at instant  $t_1$ .
- Next steps continue iterations 1-3. At the end of the process we will get the external estimate  $E(a(t), Q(t))$  of the tube  $X(t)$  with accuracy tending to zero when  $m \rightarrow \infty$ .

Consider the estimation of the viable trajectory tube  $X(t)$  of the system (29) under constraint (6). We modify Algorithm 1 taking into account the viability constraint (6) where we take  $Y(t) = Y = E(y_0, D)$ .

In this case from Theorems 2-4 we have the main inclusion

$$X(t_0 + \sigma, t_0, X_0) \subseteq E(a(\sigma), Q(\sigma)) \cap E(y_0, D) +$$

$$+ o(\sigma)B(0, 1), \quad X_0 = E(a, k^2 B^{-1}), \quad (35)$$

which allows to formulate the modified algorithm which is more complicated now (all notations in (35) are taken from Theorem 4).

**Algorithm 2.** Subdivide the time segment  $[t_0, T]$  into subsegments  $[t_i, t_{i+1}]$  where  $t_i = t_0 + ih$  ( $i = 1, \dots, m$ ),  $h = (T - t_0)/m$ ,  $t_m = T$ .

- Given  $X_0 = E(a, k_0^2 B^{-1})$  with  $k_0 \neq 0$ , define  $X_1 = E(a_1, Q_1)$  from (35) (as in Algorithm 1) for  $a_1 = a(\sigma)$ ,  $Q_1 = Q(\sigma)$ ,  $\sigma = h$ .
- Consider the intersection of ellipsoids  $X_1 = E(a_1, Q_1)$  and  $Y(t) = Y = E(y_0, D)$  and find the smallest (with respect to some criterion, e.g. as in [5]) ellipsoid  $X_1^* = E(a_1^*, Q_1^*)$  such that

$$E(a_1, Q_1) \cap E(y_0, D) \subset E(a_1^*, Q_1^*).$$

- Find the smallest constant  $k_1$  such that

$$E(a_1^*, Q_1^*) \subset \tilde{X}_1 = E(a_1^*, k_1^2 B^{-1}),$$

$k_1^2$  is the maximal eigenvalue of the matrix  $B^{1/2} Q_1^* B^{1/2}$ .

- Consider the system on the next interval  $[t_1, t_2]$  with  $E(a_1^*, k_1^2 B^{-1})$  as the initial ellipsoid taken at initial instant  $t_1$ .
- Next steps continue iterations 1-4. At the end of the process we will get the external estimating tube  $E(a^*(t), Q^*(t))$  of the tube  $X(t)$  with accuracy tending to zero when  $m \rightarrow \infty$ .

**Example 2.** Consider the following system

$$\begin{cases} \dot{x}_1 &= -x_1, \\ \dot{x}_2 &= \frac{1}{2}x_2 + 3(\frac{x_1^2}{4} + x_2^2), \end{cases} \quad (36)$$

$$X_0 = E(0, Q_0), \quad Q_0 = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}. \quad (37)$$

Here  $d = \{0, 3\}$ ,  $\tilde{f}(x) = x'Bx$  with  $B = Q_0^{-1}$ ,  $T = 0.15$ . Results of computer simulations based on Theorem 4 are shown at Fig. 2.

Assume now that the state constraint (6) is also present for the system (36) with  $Y = B(0, r)$ ,  $r = 2.4$ .

Applying Algorithm 2 we discover that the viability constraint becomes important in estimation only after 10th iteration so we may use there the simpler Algorithm 1. After that beginning with the 11th iteration the whole four-steps procedure works. Fig. 3-4 illustrate this estimation process.

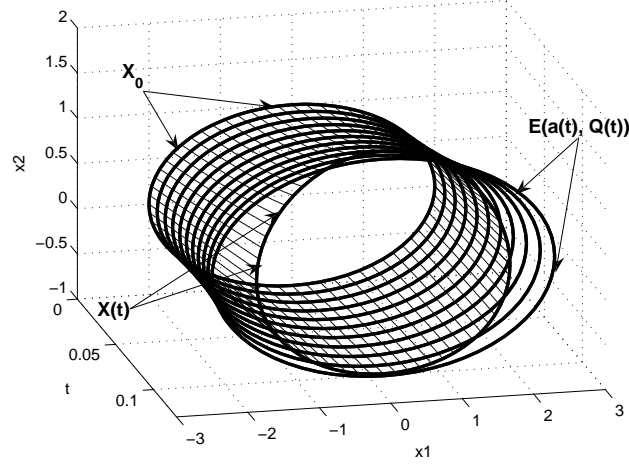


Figure 2: Trajectory tube  $X(t, t_0, X_0)$  and its external ellipsoidal tube  $E(a(t), Q(t))$ .

## 5 Internal Ellipsoidal Estimates

Consider now the internal set-valued estimates of reachable sets  $X(t)$  of the uncertain nonlinear system (29). As in the previous section we formulate first the following auxiliary result.

**Theorem 5** *Let  $X_0 = E(a, k^2 B^{-1})$  with  $k \neq 0$ . Then for the trajectory tube  $X(t)$  of the system (29) and for all  $\sigma > 0$  the following inclusion holds*

$$E(a^-(\sigma), Q^-(\sigma)) \subseteq X(t_0 + \sigma) + o(\sigma)B(0, 1), \quad \lim_{\sigma \rightarrow +0} \sigma^{-1} o(\sigma) = 0, \quad (38)$$

where

$$\begin{aligned} a^-(\sigma) &= a(\sigma) + \sigma \hat{a}, \\ Q^-(\sigma) &= Q(\sigma) + \sigma^2 \hat{Q} + 2\sigma Q(\sigma)^{1/2} (Q(\sigma)^{-1/2} \hat{Q} Q(\sigma)^{-1/2})^{1/2} Q(\sigma)^{1/2}, \end{aligned} \quad (39)$$

and  $a(\sigma)$ ,  $Q(\sigma)$  are defined in Theorem 4.

*Proof.* The proof of this result is similar to the proof of Theorem 3 [10].

Based on this result we may formulate the following scheme that gives the internal estimate of trajectory tube  $X(t)$  of the system (29).

**Algorithm 3.** Subdivide the time segment  $[t_0, T]$  into subsegments  $[t_i, t_{i+1}]$  where  $t_i = t_0 + ih$  ( $i = 1, \dots, m$ ),  $h = (T - t_0)/m$ ,  $t_m = T$ .

- Given  $X_0 = E(a, k_0^2 B^{-1})$  with  $k_0 \neq 0$ , define  $X_1 = E(a_1, Q_1)$  from Theorem 5 for  $a_1 = a^-(\sigma)$ ,  $Q_1 = Q^-(\sigma)$ ,  $\sigma = h$ .

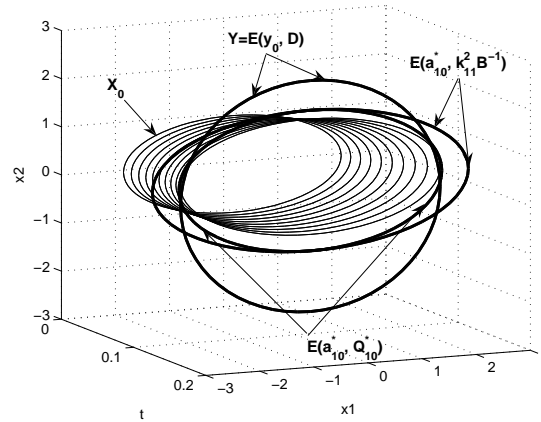


Figure 3: Viable trajectory tube  $X(t, t_0, X_0)$  and its external ellipsoidal tube  $E(a^*(t), Q^*(t))$ .

- Define  $k_1^2$  as the minimal eigenvalue of the matrix  $B^{1/2}Q_1B^{1/2}$ .
- Consider the system on the next subsegment  $[t_1, t_2]$  with  $E(a_1, k_1^2 B^{-1})$  as the initial ellipsoid at instant  $t_1$ .
- Repeat consequently the steps, at the end of the process we will get the internal estimate  $E(a(t), Q(t))$  of the tube  $X(t)$  with accuracy tending to zero when  $m \rightarrow \infty$ .

Note that  $k_1$  defined at second step of the algorithm is the largest positive constant such that  $E(a_1, k_1^2 B^{-1}) \subset E(a_1, Q_1)$ .

**Example 3.** Consider the following system

$$\begin{cases} \dot{x}_1 &= x_1 + x_1^2 + x_2^2 + u_1, \\ \dot{x}_2 &= -x_2 + u_2, \end{cases} \quad , \quad 0 \leq t \leq T. \quad (40)$$

Here  $t_0 = 0$ ,  $T = 0.3$ ,  $h = 0.025$ ,  $a_0 = \hat{a} = (0, 0)$ ,  $Q_0 = \hat{Q} = I$ . Results of computer simulations based on Theorem 5 are shown at Fig. 5.

## 6 Conclusions

The paper deals with the problems of control and state estimation for a dynamical control system described by differential inclusions with unknown but bounded initial state.

The solution to the differential system is studied through the techniques of trajectory tubes with their cross-sections  $X(t)$  being the reachable sets at instant  $t$  to control system.

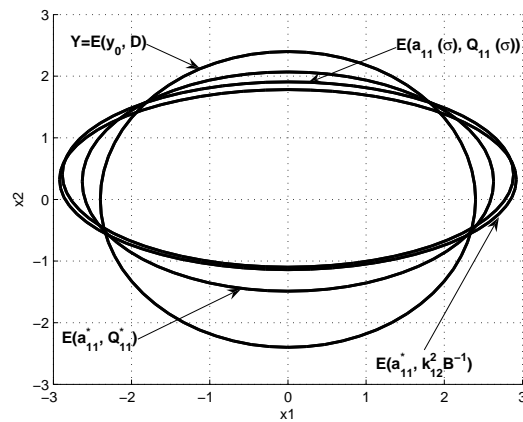


Figure 4: Steps 1-4 of 11th iteration of Algorithm 2.

Basing on the results of ellipsoidal calculus developed for linear uncertain systems we present the modified state estimation approaches which use the special nonlinear structure of the control system and simplify calculations.

Examples and numerical results related to procedures of set-valued approximations of trajectory tubes and reachable sets were also presented.

## References

- [1] J.-P. Aubin, *Viability theory*, Birkhauser, Boston, 1991.
- [2] J.-P. Aubin and H. Frankowska, *Set-valued analysis*, Birkhauser, Boston, 1990.
- [3] E.A. Barbashin, On the theory of generalized dynamic systems, *Uchen. Zap. Moscow Univ., Matematika*, 135, 110-133 (1949).
- [4] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions*, Lect. Notes in Math. , 580, 1977.
- [5] F.L. Chernousko, *State Estimation for Dynamic Systems*, CRC Press, Boca Raton, 1994.
- [6] A.L. Dontchev and E.M. Farkhi, Error estimates for discretized differential inclusions, *Computing*, 4, 349–358 (1989).
- [7] A.L. Dontchev and F. Lempio, Difference methods for differential inclusions: a survey, *SIAM Review* , 34, 263–294 (1992).

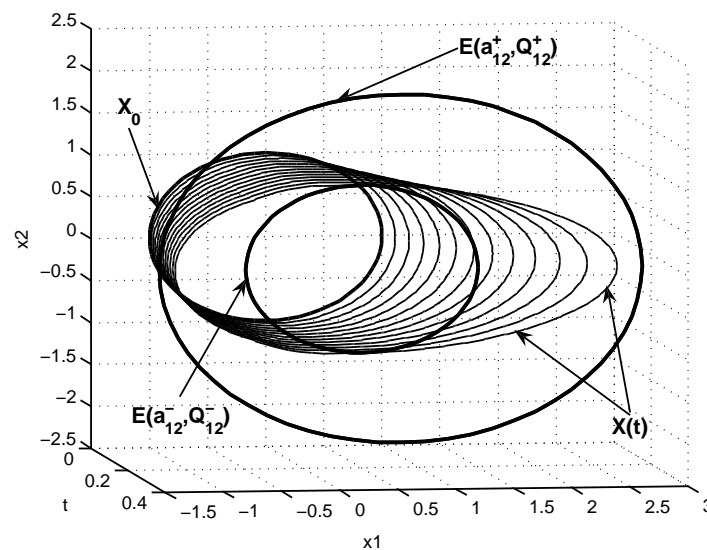


Figure 5: Trajectory tube  $X(t)$  and its external and internal ellipsoidal estimates  $E(a^+(t), Q^+(t))$ ,  $E(a^-(t), Q^-(t))$ .

- [8] A.F. Filippov, *Differential equations with discontinuous right-hand side*, Nauka, Moscow, 1985.
- [9] T.F. Filippova, A note on the evolution property of the assembly of viable solution to a differential inclusion, *Computers Math. Applic.*, 25, 115-121 (1993).
- [10] T.F. Filippova and E.V. Berezina, On State Estimation Approaches for Uncertain Dynamical Systems with Quadratic Nonlinearity: Theory and Computer Simulations, *Lecture Notes in Computer Science*, Springer, 4818, 326-333 (2008).
- [11] E.K. Kostousova and A.B. Kurzhanski, Theoretical Framework and Approximation Techniques for Parallel Computation in Set-membership State Estimation, in *Proc. of the Symposium on Modelling Analysis and Simulation, Lille, France, July 9-12, 1996*, vol. 2, 1996, 849-854.
- [12] N.N. Krasovskii, *The control of a dynamic system*, Nauka, Moscow, 1986.
- [13] N.N. Krasovskii and A.I. Subbotin, *Positional differential games*, Springer-Verlag, 1988.
- [14] A.B. Kurzhanski, *Control and observation under conditions of uncertainty*, Nauka, Moscow, 1977.



- [15] A.B. Kurzhanski and T.F. Filippova, On the theory of trajectory tubes — a mathematical formalism for uncertain dynamics, viability and control, in *Advances in Nonlinear Dynamics and Control: a Report from Russia* (A.B. Kurzhanski, ed.), Progress in Systems and Control Theory, Birkhauser, 1993, 17, 122-188.
- [16] A.B. Kurzhanski and O.I. Nikonov, On the control strategy synthesis problem. Evolution equations and set-valued integration, *Doklady Akad. Nauk SSSR*, 311, 788-793 (1990).
- [17] A.B. Kurzhanski and I. Valyi, *Ellipsoidal Calculus for Estimation and Control*, Birkhauser, Boston, 1997.
- [18] A.B. Kurzhanski and V.M. Veliov, *Set-valued Analysis and Differential Inclusions*, Progress in Systems and Control Theory, Birkhauser, Boston, 1990.
- [19] A.I. Panasyuk, Equations of attainable set dynamics, Part 1: Integral funnel equations, *J. Optimiz. Theory Appl.*, 2, 349–366 (1990).
- [20] E. Roxin, On the generalized dynamical systems defined by contingent equations, *J. of Diff. Equations*, 1, 188-205 (1965).
- [21] E. Walter and L. Pronzato, *Identification of parametric models from experimental data*, Communications and Control Engineering Series, Springer, London, 1997.

# Approximate formulae for fractional derivatives by means of Sinc methods

Tomoaki Okayama<sup>†</sup>, Takayasu Matsuo, Masaaki Sugihara

Graduate School of Information Science and Technology,

The University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan

<sup>†</sup> Tomoaki.Okayama@mist.i.u-tokyo.ac.jp

## Abstract

In this paper, two new approximate formulae for fractional derivatives are developed by means of Sinc methods. The difference of the two formulae is the variable transformations incorporated; the single exponential transformation and the double exponential transformation. We give error analysis of the formulae, and show that these formulae archive exponential convergence. Numerical examples that confirm the analysis are also given.

Keywords: fractional derivative, numerical approximation, Sinc methods

## 1 Introduction

In the last few decades, mathematical models with fractional derivatives have been used in the fields of physics [6], engineering [16], chemistry [14], biology [9], control theory [3], and many others [1, 2, 7]. We consider two types of derivatives of order  $p$ : Riemann–Liouville type ( $\mathbf{D}_a^p f$ ) and Caputo’s type ( $\mathbf{D}_a^p f$ ), which are defined by

$$\mathbf{D}_a^p[f](t) = \left(\frac{d}{dt}\right)^{\lfloor p \rfloor + 1} \left[ \mathbf{I}_a^{\lfloor p \rfloor - p + 1} f \right](t), \quad t > a, \quad (1)$$

$$\mathbf{D}_a^p[f](t) = \mathbf{I}_a^{\lfloor p \rfloor - p + 1} \left[ \left(\frac{d}{dt}\right)^{\lfloor p \rfloor + 1} f \right](t), \quad t > a, \quad (2)$$

## 1 INTRODUCTION

respectively, where  $\mathbf{I}_a^q f$  is the *Riemann–Liouville fractional integral* of order  $q$ ,

$$\mathbf{I}_a^q[f](t) = \frac{1}{\Gamma(q)} \int_a^t \frac{f(s) \, ds}{(t-s)^{1-q}}, \quad t > a. \quad (3)$$

In what follows, we assume  $p, q \in (0, 1)$ . In this case, approximating fractional derivatives with high accuracy is not an easy task, because there is a weakly singular kernel called the *Abel kernel* in (3). Typical numerical methods for fractional derivatives in the literature are reviewed in some books cited above and some papers [4, 5]. The convergence rates of those methods are all of polynomial:  $O(n^{-\gamma})$ , where  $n$  denotes the number of evaluation of  $f$ , and  $\gamma$  is a positive constant.

Recently, an “exponentially” converging approximate formula based on Chebyshev polynomials has been proposed by Sugiura–Hasegawa [19]. In their beautiful work, they have extended the so-called Clenshaw–Curtis rule for the definite integral  $\int_{-1}^1 f(s) \, ds$  to the fractional derivative of Caputo’s type (2), and also pointed out that the formula is also applicable to the Riemann–Liouville type (1) through the relation  $\mathbf{D}_a^p[f](t) = D_a^p[f](t) + f(a)(t-a)^{-p}/\Gamma(1-p)$ . They have shown that the formula converges uniformly on the given interval  $[a, b]$  with the exponential rate,  $O(e^{-\gamma^n})$ , under the assumption that  $f$  is analytic on an elliptic domain that contains the interval  $[a, b]$ . In general, however,  $f$  does not satisfy this assumption. In fact, the solution of fractional differential equations may have a singularity at the endpoint,  $t = a$ , due to the Abel kernel [8]. In such cases, their formula loses the fast convergence.

On the other hand, for such singular functions, it is known in the wide range of numerical analysis that Sinc methods are quite effective (see, for example, Stenger [17]). In fact, Riley [15] employed techniques in Sinc methods to approximate integrals of the form (3), and obtained exponential convergence,  $O(e^{-\gamma \sqrt{n}})$ , despite singularities in the kernel and the function  $f$ . This result has then been extended by Mori et al. [10] and the present authors [13], and it turned out that the convergence rate of the method can be improved to  $O(e^{-\gamma n / \log n})$ . The key in this improvement is the replacement of

Tomoaki Okayama et al.

the variable transformation; the standard *Single Exponential (SE) transformation* employed in Riley's method was replaced with a stronger transformation, the so-called *Double Exponential (DE) transformation* [11, 18]. The latter methods, i.e. the Sinc methods incorporated with the DE transformation, are called *DE-Sinc methods*, while the former ones are referred to *SE-Sinc methods*, accordingly.

As a natural extension of these results, in the present paper we propose two new approximate formulae for Caputo's fractional derivative (2); either based on the SE-Sinc and DE-Sinc methods. It is then shown theoretically and numerically that the convergence rate is  $O(e^{-\gamma\sqrt{n}})$  in the first formula, and  $O(e^{-\gamma n/\log n})$  in the second formula. These formulae are also applicable to the Riemann–Liouville fractional derivative (1) in the same manner as in Sugiura–Hasegawa [19].

This paper is organized as follows. The main results are stated in Section 2. In Section 3, we show numerical examples of the new formulae, and compare them with the one by Sugiura–Hasegawa. The proofs of the main theorems are given in Section 4.

## 2 Approximate formulae and their error analysis

The main tool to derive approximate formulae is the *Sinc approximation*:

$$F(\tau) \approx \sum_{j=-N}^N F(jh)S(j, h)(\tau), \quad \tau \in \mathbb{R}, \quad (4)$$

where  $S(j, h)(\tau)$  is the *Sinc function* defined by  $S(j, h)(\tau) = \sin\{\pi(\tau/h - j)\}/\{\pi(\tau/h - j)\}$ .

The so-called *Sinc quadrature* rule is derived by integrating the both sides of (4):

$$\int_{-\infty}^{\infty} F(\tau) d\tau \approx \sum_{j=-N}^N F(jh) \int_{-\infty}^{\infty} S(j, h)(\tau) d\tau = h \sum_{j=-N}^N F(jh). \quad (5)$$

Note that the variable  $\tau$  in these formulae moves on the whole real line. If the function to be approximated is defined on a finite domain, variable transformation should be

## 2 APPROXIMATE FORMULAE AND THEIR ERROR ANALYSIS

employed in (4) or (5). There are two transformations, the SE transformation and the DE transformation, which are defined by

$$\begin{aligned} t = \psi_{a,b}^{\text{SE}}(\tau) &= \frac{b-a}{2} \tanh\left(\frac{\tau}{2}\right) + \frac{b+a}{2}, \\ t = \psi_{a,b}^{\text{DE}}(\tau) &= \frac{b-a}{2} \tanh\left(\frac{\pi}{2} \sinh(\tau)\right) + \frac{b+a}{2}. \end{aligned}$$

Both transformations map  $\tau \in \mathbb{R}$  onto  $t \in (a, b)$ . Their inverse functions are:

$$\begin{aligned} \tau = \{\psi_{a,b}^{\text{SE}}\}^{-1}(t) &= \log\left(\frac{t-a}{b-t}\right), \\ \tau = \{\psi_{a,b}^{\text{DE}}\}^{-1}(t) &= \log\left[\frac{1}{\pi} \log\left(\frac{t-a}{b-t}\right) + \sqrt{1 + \left\{\frac{1}{\pi} \log\left(\frac{t-a}{b-t}\right)\right\}^2}\right]. \end{aligned}$$

### 2.1 Derivation of a formula by means of the SE-Sinc methods

Recall that Caputo's fractional derivative is defined by  $D_a^p[f](t) = \mathbf{I}_a^{1-p}[f'](t)$ . Our basic idea is to approximate the integral part ( $\mathbf{I}_a^{1-p}$ ) based on the idea in Riley [15], and the derivative part ( $\frac{d}{dt}$ ) based on the idea in Stenger [17], respectively. Finally we combine them to approximate the target:  $D_a^p f$ .

First we consider the approximation of  $\mathbf{I}_a^{1-p} g$  for a given function  $g$ . Changing the original integral interval  $(a, t)$  to  $\mathbb{R}$  by the variable transformation  $s = \psi_{a,t}^{\text{SE}}(\sigma)$ , we have

$$\mathbf{I}_a^{1-p}[g](t) = \int_{-\infty}^{\infty} \frac{g(\psi_{a,t}^{\text{SE}}(\sigma)) \{\psi_{a,t}^{\text{SE}}\}'(\sigma) d\sigma}{\Gamma(1-p)(t - \psi_{a,t}^{\text{SE}}(\sigma))^p} = \frac{(t-a)^{1-p}}{\Gamma(1-p)} \int_{-\infty}^{\infty} \frac{g(\psi_{a,t}^{\text{SE}}(\sigma)) d\sigma}{(1 + e^{-\sigma})(1 + e^{\sigma})^{1-p}}.$$

Note that the weakly singular integrand (the Abel kernel) is translated to a smooth function. Applying the quadrature rule (5) to the translated integral, we obtain the approximate formula for the integral part:

$$\mathbf{I}_a^{1-p}[g](t) \approx \mathcal{I}_N^{\text{SE}}[g](t) = \frac{(t-a)^{1-p}}{\Gamma(1-p)} h \sum_{k=-N}^N \frac{g(\psi_{a,t}^{\text{SE}}(kh))}{(1 + e^{-kh})(1 + e^{kh})^{1-p}}. \quad (6)$$

Tomoaki Okayama et al.

Here  $h$  is a mesh size suitably chosen depending on  $N$ , which will be described later.

Next we consider the approximation of  $f'$ . Let us define a function  $Q_{a,b}$  as  $Q_{a,b}(t) = (t-a)(b-t)$ . Putting  $F(\tau) = f(\psi_{a,b}^{\text{SE}}(\tau))/Q_{a,b}(\psi_{a,b}^{\text{SE}}(\tau))$  in (4), we have

$$\frac{f(\psi_{a,b}^{\text{SE}}(\tau))}{Q_{a,b}(\psi_{a,b}^{\text{SE}}(\tau))} \approx \sum_{j=-N}^N \frac{f(\psi_{a,b}^{\text{SE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{SE}}(jh))} S(j, h)(\tau), \quad \tau \in \mathbb{R},$$

which is equivalent to:

$$f(t) \approx C_N^{\text{SE}}[f](t) = \sum_{j=-N}^N \frac{f(\psi_{a,b}^{\text{SE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{SE}}(jh))} Q_{a,b}(t) S(j, h)(\{\psi_{a,b}^{\text{SE}}\}^{-1}(t)), \quad t \in (a, b). \quad (7)$$

Differentiating the both sides gives an approximate formula for  $f'$ , i.e.  $f' \approx \{C_N^{\text{SE}}f\}'$ .

Using this and (6) with  $g = f'$ , we finally obtain the desired formula as follows:

$$D_a^p[f](t) = \mathbf{I}_a^{1-p}[f'](t) \approx \mathcal{I}_N^{\text{SE}}[f'](t) \approx \mathcal{I}_N^{\text{SE}}[\{C_N^{\text{SE}}f\}'](t). \quad (8)$$

## 2.2 Derivation of a formula by means of the DE-Sinc methods

We consider the use of the DE transformation instead of the SE transformation here.

For the integral part  $\mathbf{I}_a^{1-p} g$ , we apply  $s = \psi_{a,t}^{\text{DE}}(\sigma)$ , then  $\mathbf{I}_a^{1-p} g$  is translated into

$$\mathbf{I}_a^{1-p}[g](t) = \frac{(t-a)^{1-p}}{\Gamma(1-p)} \int_{-\infty}^{\infty} \frac{\pi \cosh(\sigma) g(\psi_{a,t}^{\text{DE}}(\sigma)) d\sigma}{(1 + e^{-\pi \sinh(\sigma)})(1 + e^{\pi \sinh(\sigma)})^{1-p}}.$$

Applying the quadrature rule (5) to this integral, we obtain the approximate formula:

$$\mathbf{I}_a^{1-p}[g](t) \approx \mathcal{I}_N^{\text{DE}}[g](t) = \frac{(t-a)^{1-p}}{\Gamma(1-p)} h \sum_{k=-N}^N \frac{\pi \cosh(kh) g(\psi_{a,t}^{\text{DE}}(kh))}{(1 + e^{-\pi \sinh(kh)})(1 + e^{\pi \sinh(kh)})^{1-p}}.$$

The derivative part can be handled in the same manner. Similar to (7), using

$$f(t) \approx C_N^{\text{DE}}[f](t) = \sum_{j=-N}^N \frac{f(\psi_{a,b}^{\text{DE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))} Q_{a,b}(t) S(j, h)(\{\psi_{a,b}^{\text{DE}}\}^{-1}(t)), \quad t \in (a, b),$$

## 2 APPROXIMATE FORMULAE AND THEIR ERROR ANALYSIS

and differentiating both sides, we have  $f' \approx \{C_N^{\text{DE}} f\}'$ . Then we obtain the formula:

$$D_a^p[f](t) = \mathbf{I}_a^{1-p}[f'](t) \approx \mathcal{I}_N^{\text{DE}}[f'](t) \approx \mathcal{I}_N^{\text{DE}}[\{C_N^{\text{DE}} f\}'](t). \quad (9)$$

### 2.3 Results of error analysis

We here state the error analysis results of the presented approximate formulae, while their proofs are left to Section 4. Let us introduce the following function space.

**Definition 1.** Let  $\mathcal{D}$  be a simply-connected domain which satisfies  $(a, b) \subset \mathcal{D}$ , and let  $\alpha$  be a positive constant. Then  $\mathbf{L}_\alpha(\mathcal{D})$  denotes the family of all functions  $f$  that are analytic on  $\mathcal{D}$ , and satisfy  $|f(z)| \leq C|Q_{a,b}^\alpha(z)|$  for a positive constant  $C$  and all  $z \in \mathcal{D}$ .

In the statement of theorems below,  $\mathcal{D}$  is either  $\psi_{a,b}^{\text{SE}}(\mathcal{D}_d)$  or  $\psi_{a,b}^{\text{DE}}(\mathcal{D}_d)$ , where

$$\begin{aligned} \psi_{a,b}^{\text{SE}}(\mathcal{D}_d) &= \left\{ z \in \mathbb{C} : \left| \arg \left( \frac{z-a}{b-z} \right) \right| < d \right\}, \\ \psi_{a,b}^{\text{DE}}(\mathcal{D}_d) &= \left\{ z \in \mathbb{C} : \left| \arg \left[ \frac{1}{\pi} \log \left( \frac{z-a}{b-z} \right) + \sqrt{1 + \left\{ \frac{1}{\pi} \log \left( \frac{z-a}{b-z} \right) \right\}^2} \right] \right| < d \right\}. \end{aligned}$$

These are domains that are mapped by the SE or DE transformation from a strip domain

$$\mathcal{D}_d = \{\zeta \in \mathbb{C} : |\text{Im } \zeta| < d\}, \quad (10)$$

for a positive constant  $d$ . With these notations, the approximate errors of the formula (8) and (9) are analyzed as follows.

**Theorem 1.** Let  $(f/Q_{a,b}) \in \mathbf{L}_\alpha(\psi_{a,b}^{\text{SE}}(\mathcal{D}_d))$  for  $d$  with  $0 < d < \pi$ . Let  $\mu = \min\{1 - p, \alpha\}$ ,  $N$  be a positive integer, and  $h$  be selected by  $h = \sqrt{\pi d / (\mu N)}$ . Then there exists a constant  $C$  independent of  $N$  such that

$$\max_{t \in [a, b]} |D_a^p[f](t) - \mathcal{I}_N^{\text{SE}}[\{C_N^{\text{SE}} f\}'](t)| \leq C N e^{-\sqrt{\pi d \mu N}}. \quad (11)$$

Tomoaki Okayama et al.

**Theorem 2.** Let  $(f/Q_{a,b}) \in \mathbf{L}_\alpha(\psi_{a,b}^{\text{DE}}(\mathcal{D}_d))$  for  $d$  with  $0 < d < \pi/2$ , and let  $f$  be analytic at  $b$ . Let  $\mu = \min\{1 - p, \alpha\}$ ,  $N$  be a positive integer with  $N > \mu/(2d)$ , and  $h$  be selected by  $h = \log(2dN/\mu)/N$ . Then there exists a constant  $C$  independent of  $N$  such that

$$\max_{t \in [a, b]} |D_a^p[f](t) - I_N^{\text{DE}}[\{C_N^{\text{DE}} f\}'](t)| \leq C \frac{N}{\log(2dN/\mu)} e^{-\pi dN/\log(2dN/\mu)}.$$

The number of evaluation of  $f$  in these approximation formulae is  $n = 2N + 1$ , which means the convergence rate is  $O(e^{-\gamma\sqrt{n}})$  in the SE-Sinc case, and  $O(e^{-\gamma n/\log n})$  in the DE-Sinc case for some  $\gamma > 0$ ; in both cases the errors decay exponentially.

**Remark 1.** The assumption  $(f/Q_{a,b}) \in \mathbf{L}_\alpha(\mathcal{D})$  may seem to be not practical since the function  $f$  must be zero at the endpoints by the condition  $|f(z)/Q_{a,b}(z)| \leq C|Q_{a,b}^\alpha(z)|$ . But actually, functions in a certain wider, and reasonable space can be translated to those satisfying the assumption (see Stenger [17, § 4]).

### 3 Numerical examples

In this section we consider two test functions,  $f_1(t) = t^{4/3}(1-t)^2/\Gamma(7/3)$  and  $f_2(t) = t^2(1-t)^2e^t$ , and their  $1/2$ -order derivatives in Caputo's sense on the interval  $(0, 1)$ :

$$\begin{aligned} D_0^{1/2}[f_1](t) &= \frac{t^{5/6}}{\Gamma(11/6)} \frac{280t^2 - 476t + 187}{187}, \\ D_0^{1/2}[f_2](t) &= \frac{1}{16} \left[ \frac{t^{1/2}}{\Gamma(3/2)} (8t^3 - 4t^2 - 22t + 31) + e^t \operatorname{erf}(\sqrt{t}) \{8t(2t^3 - 7t + 8) - 31\} \right]. \end{aligned}$$

Let  $\pi_m$  denote an arbitrary positive number less than  $\pi$ . Then the function  $f_1$  satisfies  $(f_1/Q_{0,1}) \in \mathbf{L}_{1/3}(\psi_{0,1}^{\text{SE}}(\mathcal{D}_{\pi_m}))$  and  $(f_1/Q_{0,1}) \in \mathbf{L}_{1/3}(\psi_{0,1}^{\text{DE}}(\mathcal{D}_{\pi_m/2}))$ , and the function  $f_2$  satisfies  $(f_2/Q_{0,1}) \in \mathbf{L}_1(\psi_{0,1}^{\text{SE}}(\mathcal{D}_{\pi_m}))$  and  $(f_2/Q_{0,1}) \in \mathbf{L}_1(\psi_{0,1}^{\text{DE}}(\mathcal{D}_{\pi_m/2}))$ . In actual computations, we set  $\pi_m = 3.14$ , and then  $h$  can be selected according to Theorem 1 or Theorem 2.

The numerical result of  $D_0^{1/2} f_1$  is shown in Fig. 1, and the one of  $D_0^{1/2} f_2$  is shown



## 4 PROOFS OF THE THEOREMS IN SECTION 2

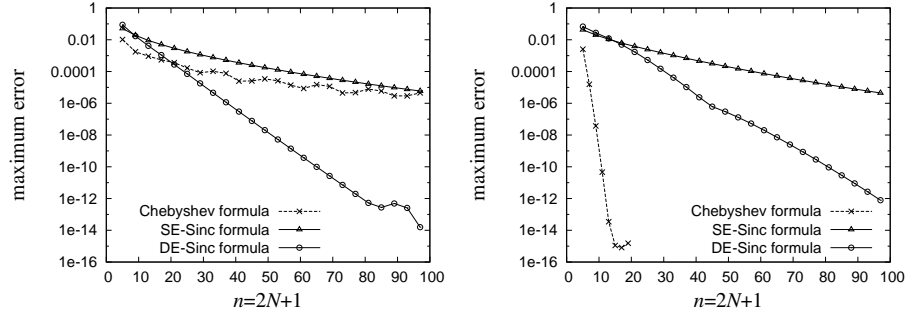


Fig. 1. Approximation errors of  $D_0^{1/2} f_1$ . Fig. 2. Approximation errors of  $D_0^{1/2} f_2$ .

in Fig. 2. Both of the computation programs are written in C with double-precision floating-point arithmetic. The errors are checked on  $t = 0.01, 0.02, \dots, 0.99$ , and the maximum error of them is plotted on the graphs. There are three plot lines in both graphs; the formula by Sugiura–Hasegawa [19] (dashed line with  $\times$  points), by the SE-Sinc methods (solid line with  $\triangle$  points), and by the DE-Sinc methods (solid line with  $\circ$  points). The convergence profiles of the Chebyshev formula are different between Fig. 1 and Fig. 2. This should be caused by the singularity of the function  $f_1$  at the endpoint,  $t = 0$ . In contrast, we can see that the results of the SE-Sinc formula and the DE-Sinc formula are consistent with Theorem 1 or Theorem 2 in both graphs.

## 4 Proofs of the theorems in Section 2

### 4.1 Proof of Theorem 1 (the SE-Sinc case)

The following two theorems are critical to prove Theorem 1.

**Theorem 3.** Let the assumptions of Theorem 1 are fulfilled. Then there exists a constant  $C$  independent of  $N$  such that

$$\max_{t \in [a, b]} \left| \mathbf{I}_a^{1-p}[f'](t) - \mathcal{I}_N^{\text{SE}}[f'](t) \right| \leq C e^{-\sqrt{\pi d \mu N}}.$$

Tomoaki Okayama et al.

**Theorem 4** (Stenger [17, Corollary of Theorem 4.4.2]). Let the assumptions of Theorem 1 are fulfilled. Then there exists a constant  $C$  independent of  $N$  such that

$$\sup_{t \in (a, b)} \left| \frac{d}{dt} \{f(t) - C_N^{\text{SE}}[f](t)\} \right| \leq C N e^{-\sqrt{\pi d \mu N}}.$$

Using these theorems and the trivial fact  $\sup_N \|I_N^{\text{SE}}\|_{C([a, b])} < \infty$ , we get (11). In what follows, we prove Theorem 3. The next theorem is the base of the error analysis.

**Theorem 5** (Stenger [17, Theorem 4.2.6]). Let  $(FQ_{a,b}) \in \mathbf{L}_\beta(\psi_{a,b}^{\text{SE}}(\mathcal{D}_d))$  for  $d$  with  $0 < d < \pi$ , let  $N$  be a positive integer, and  $h$  be selected by  $h = \sqrt{2\pi d/(\beta N)}$ . Then there exists a constant  $C$  independent of  $N$  such that

$$\left| \int_a^b F(s) ds - h \sum_{k=-N}^N F(\psi_{a,b}^{\text{SE}}(kh)) \{\psi_{a,b}^{\text{SE}}\}'(kh) \right| \leq C e^{-\sqrt{2\pi d \beta N}}.$$

Let us apply this theorem to the approximation (6). If we put  $F(s) = g(s)/(t-s)^p$  in this theorem, and if  $g$  is analytic and bounded uniformly on  $\psi_{a,t}^{\text{SE}}(\mathcal{D}_d)$  for all  $t \in [a, b]$ , then  $(FQ_{a,t}) \in \mathbf{L}_{1-p}(\psi_{a,t}^{\text{SE}}(\mathcal{D}_d))$ . Furthermore if we set  $\mu = \min\{1-p, \alpha\}$ , then  $(FQ_{a,t}) \in \mathbf{L}_\mu(\psi_{a,t}^{\text{SE}}(\mathcal{D}_d))$  since clearly  $\mathbf{L}_\nu(\psi_{a,t}^{\text{SE}}(\mathcal{D}_d)) \subseteq \mathbf{L}_\rho(\psi_{a,t}^{\text{SE}}(\mathcal{D}_d))$  if  $\nu \geq \rho$ . Therefore we obtain the next result.

**Lemma 1.** Assume that there exists a constant  $d$  with  $0 < d < \pi$  such that  $g$  is analytic and bounded uniformly on  $\psi_{a,t}^{\text{SE}}(\mathcal{D}_d)$  for all  $t \in [a, b]$ . Let  $\mu = \min\{1-p, \alpha\}$ ,  $N$  be a positive integer, and  $h$  be selected by  $h = \sqrt{2\pi d/(\mu N)}$ . Then there exists a constant  $C$  independent of  $N$  such that

$$\max_{t \in [a, b]} \left| \mathbf{I}_a^{1-p}[g](t) - \mathcal{I}_N^{\text{SE}}[g](t) \right| \leq C e^{-\sqrt{2\pi d \mu N}}. \quad (12)$$

We can relax the condition on  $g$  using the following lemma.

**Lemma 2.** Let  $g$  be analytic and bounded on  $\psi_{a,b}^{\text{SE}}(\mathcal{D}_d)$  for  $d$  with  $0 < d < \pi$ . Then  $g$  is analytic and bounded uniformly on  $\psi_{a,t}^{\text{SE}}(\mathcal{D}_d)$  for all  $t \in [a, b]$ .

## 4 PROOFS OF THE THEOREMS IN SECTION 2

*Proof.* We shall establish this lemma if we prove for all  $t \in [a, b]$  that  $\psi_{a,t}^{\text{SE}}(\mathcal{D}_d) \subseteq \psi_{a,b}^{\text{SE}}(\mathcal{D}_d)$ , which is equivalent to “ $|\text{Im}\{\psi_{a,t}^{\text{SE}}\}^{-1}(z)| < d \Rightarrow |\text{Im}\{\psi_{a,b}^{\text{SE}}\}^{-1}(z)| < d$ ” (recall that  $\mathcal{D}_d$  is defined by (10)). Set  $\zeta(t) = \{\psi_{a,t}^{\text{SE}}\}^{-1}(z)$  for simplicity. It is sufficient to show that  $|\text{Im}\zeta(t)|$  is a monotonically decreasing function, since from this we have  $|\text{Im}\zeta(b)| \leq |\text{Im}\zeta(t)| < d$ . Let  $x, y \in \mathbb{R}$  and set  $z = x + iy$ . Then  $\text{Im}\zeta(t)$  is expressed as

$$\text{Im}\zeta(t) = \arg\left(\frac{z-a}{t-z}\right) = \arg\left(\frac{ax+tx-at-x^2-y^2}{(t-x)^2+y^2} + i\frac{(t-a)y}{(t-x)^2+y^2}\right).$$

Considering  $\cos(\text{Im}\zeta(t))$  and its derivative, we have

$$\begin{aligned} \cos(\text{Im}\zeta(t)) &= \frac{ax+tx-at-x^2-y^2}{\sqrt{(ax+tx-at-x^2-y^2)^2 + (t-a)^2y^2}}, \\ \frac{d}{dt}\cos(\text{Im}\zeta(t)) &= \frac{(t-a)((a-x)^2+y^2)y^2}{\{((a-x)^2+y^2)((t-x)^2+y^2)\}^{3/2}} \geq 0. \end{aligned}$$

Thus  $\cos(\text{Im}\zeta(t))$  is a monotonically increasing function. Since  $-\pi < \text{Im}\zeta(t) \leq \pi$  and  $\cos(-\text{Im}\zeta(t)) = \cos(\text{Im}\zeta(t))$ , we can see that  $|\text{Im}\zeta(t)|$  is monotonically decreasing. ■

Therefore Lemma 1 can be rewritten as follows.

**Lemma 3.** Let  $g$  be analytic and bounded on  $\psi_{a,b}^{\text{SE}}(\mathcal{D}_d)$  for  $d$  with  $0 < d < \pi$ . Let  $\mu = \min\{1-p, \alpha\}$ ,  $N$  be a positive integer, and  $h$  be selected by  $h = \sqrt{2\pi d/(\mu N)}$ . Then there exists a constant  $C$  independent of  $N$  such that (12) holds.

If  $(f/Q_{a,b}) \in \mathbf{L}_\alpha(\psi_{a,b}^{\text{SE}}(\mathcal{D}_d))$  (assumption in Theorem 1) holds, then  $f'$  is analytic and bounded on  $\psi_{a,b}^{\text{SE}}(\mathcal{D}_{d-\epsilon})$  for any  $\epsilon$  with  $0 < \epsilon < d$ . Choosing  $\epsilon = d/2$  and using Lemma 3, we obtain Theorem 3.

## 4.2 Proof of Theorem 2 (the DE-Sinc case)

Since  $\sup_N \|I_N^{\text{DE}}\|_{C([a,b])} < \infty$ , Theorem 2 can be proved in a similar way to the SE-Sinc case, by showing the following two theorems.

Tomoaki Okayama et al.

**Theorem 6.** Let the assumptions of Theorem 2 are fulfilled. Then there exists a constant  $C$  independent of  $N$  such that

$$\max_{t \in [a, b]} \left| \mathbf{I}_a^{1-p} [f'](t) - \mathcal{I}_N^{\text{DE}} [f'](t) \right| \leq C e^{-\pi d N / \log(2dN/\mu)}.$$

**Theorem 7.** Let the assumptions of Theorem 2 are fulfilled. Then there exists a constant  $C$  independent of  $N$  such that

$$\sup_{t \in (a, b)} \left| \frac{d}{dt} \{f(t) - C_N^{\text{DE}} [f](t)\} \right| \leq C \frac{N}{\log(2dN/\mu)} e^{-\pi d N / \log(2dN/\mu)}.$$

We first give the proof of Theorem 7, which is relatively short.

#### 4.2.1 Proof of Theorem 7 (approximation error of derivatives)

We easily obtain that

$$\begin{aligned} \left| \frac{d}{dt} \{f(t) - C_N^{\text{DE}} [f](t)\} \right| &\leq \left| \frac{d}{dt} \left\{ f(t) - \sum_{j=-\infty}^{\infty} \frac{f(\psi_{a,b}^{\text{DE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))} Q_{a,b}(t) S(j, h) (\{\psi_{a,b}^{\text{DE}}\}^{-1}(t)) \right\} \right| \\ &\quad + \sum_{|j| > N} \left| \frac{f(\psi_{a,b}^{\text{DE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))} \right| \left| \frac{d}{dt} \{Q_{a,b}(t) S(j, h) (\{\psi_{a,b}^{\text{DE}}\}^{-1}(t))\} \right|. \quad (13) \end{aligned}$$

Let us examine the first term. We need the following definition for it.

**Definition 2.** Let  $\mathcal{D}_d(\epsilon)$  be defined for  $0 < \epsilon < 1$  by  $\mathcal{D}_d(\epsilon) = \{\zeta \in \mathbb{C} : |\operatorname{Re} \zeta| < 1/\epsilon, |\operatorname{Im} \zeta| < d(1 - \epsilon)\}$ . Then  $\mathbf{H}^1(\mathcal{D}_d)$  denotes the family of all functions  $F$  that are analytic on  $\mathcal{D}_d$ , and such that  $\mathcal{N}_1(F, d) = \lim_{\epsilon \rightarrow 0} \oint_{\partial \mathcal{D}_d(\epsilon)} |F(\zeta)| |d\zeta| < \infty$ .

Then the next assertion holds for any conformal map  $\psi$  that satisfies  $\psi(\mathbb{R}) = (a, b)$ .

**Theorem 8** (Stenger [17, part of Theorem 4.4.2]). Assume the next two conditions:

$$(A1) \quad f(\psi(\cdot))/Q_{a,b}(\psi(\cdot)) \in \mathbf{H}^1(\mathcal{D}_d),$$

$$(A2) \quad \sup_{t \in (a, b), -\pi/h \leq s \leq \pi/h} \left| \frac{d}{dt} (Q_{a,b}(t) e^{is\psi^{-1}(t)}) \right| \leq C/h \text{ with } C \text{ depending only on } \psi \text{ and } Q_{a,b}.$$

## 4 PROOFS OF THE THEOREMS IN SECTION 2

Then there exists a constant  $\tilde{C}$ , depending only on  $\psi$ ,  $Q$ ,  $d$  and  $f$ , such that

$$\sup_{t \in (a, b)} \left| \frac{d}{dt} \left\{ f(t) - \sum_{j=-\infty}^{\infty} \frac{f(\psi(jh))}{Q_{a,b}(\psi(jh))} Q_{a,b}(t) S(j, h)(\psi^{-1}(t)) \right\} \right| \leq \tilde{C} \frac{e^{-\pi d/h}}{h}.$$

We show that (A1) and (A2) are fulfilled with  $\psi(t) = \psi_{a,b}^{\text{DE}}(t)$  under the assumption that  $(f/Q_{a,b}) \in \mathbf{L}_{\alpha}(\psi_{a,b}^{\text{DE}}(\mathcal{D}_d))$ . For (A1), it is sufficient to prove  $\mathcal{N}_1(Q_{a,b}^{\alpha}(\psi_{a,b}^{\text{DE}}(\cdot)), d)$  is finite, since  $|f(z)/Q_{a,b}(z)| \leq C|Q_{a,b}^{\alpha}(z)|$  holds by the assumption (recall Definition 1). The next lemma shows the desired claim.

**Lemma 4** (Okayama et al. [12, Lemma 4.6]). Let  $\alpha$  and  $d$  be positive constants. Then  $\mathcal{N}_1(Q_{a,b}^{\alpha}(\psi_{a,b}^{\text{DE}}(\cdot)), d)$  is finite for any  $d \in (0, \pi/2)$ .

Using the Leibniz rule and the following inequality:

$$\frac{Q_{a,b}(t)}{\{\psi_{a,b}^{\text{DE}}\}'(\{\psi_{a,b}^{\text{DE}}\}^{-1}(t))} = \frac{(t-a)(b-t)}{\frac{\pi(t-a)(b-t)}{b-a} \sqrt{1 + \left\{ \frac{1}{\pi} \log \left( \frac{t-a}{b-t} \right) \right\}^2}} \leq \frac{b-a}{\pi},$$

we easily show the condition (A2).

**Lemma 5.** The condition (A2) in Theorem 8 holds with  $\psi(t) = \psi_{a,b}^{\text{DE}}(t)$ .

Therefore we can use Theorem 8 to evaluate the first term in (13) as follows.

**Lemma 6.** Let  $(f/Q_{a,b}) \in \mathbf{L}_{\alpha}(\psi_{a,b}^{\text{DE}}(\mathcal{D}_d))$  for  $d$  with  $0 < d < \pi/2$ . Then there exists a constant  $C$  independent of  $h$  such that

$$\sup_{t \in (a, b)} \left| \frac{d}{dt} \left\{ f(t) - \sum_{j=-\infty}^{\infty} \frac{f(\psi_{a,b}^{\text{DE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))} Q_{a,b}(t) S(j, h)(\{\psi_{a,b}^{\text{DE}}\}^{-1}(t)) \right\} \right| \leq C \frac{e^{-\pi d/h}}{h}.$$

There remains to evaluate the second term in (13); this is done by the next lemma.

**Lemma 7.** Let  $(f/Q_{a,b}) \in \mathbf{L}_{\alpha}(\psi_{a,b}^{\text{DE}}(\mathcal{D}_d))$  for  $d$  with  $0 < d < \pi/2$ . Then there exists

Tomoaki Okayama et al.

a constant  $C$  independent of  $h$  and  $N$  such that

$$\sup_{t \in (a, b)} \sum_{|j| > N} \left| \frac{f(\psi_{a,b}^{\text{DE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))} \right| \left| \frac{d}{dt} \left\{ Q_{a,b}(t) S(j, h) (\{\psi_{a,b}^{\text{DE}}\}^{-1}(t)) \right\} \right| \leq C \frac{1}{h^2 e^{Nh}} e^{-\frac{\pi}{2} \alpha \exp(Nh)}. \quad (14)$$

*Proof.* First, by the identity

$$Q_{a,b}(t) S(j, h) (\{\psi_{a,b}^{\text{DE}}\}^{-1}(t)) = \frac{h Q_{a,b}(t)}{2\pi} \int_{-\pi/h}^{\pi/h} e^{is[\{\psi_{a,b}^{\text{DE}}\}^{-1}(t) - jh]} ds$$

and Lemma 5, it follows that for a constant  $C_1$

$$\sup_{t \in (a, b)} \left| \frac{d}{dt} \left\{ Q_{a,b}(t) S(j, h) (\{\psi_{a,b}^{\text{DE}}\}^{-1}(t)) \right\} \right| \leq C_1/h. \quad (15)$$

Second, by the assumption  $(f/Q_{a,b}) \in \mathbf{L}_\alpha(\psi_{a,b}^{\text{DE}}(\mathcal{D}_d))$ , there exists a constant  $\tilde{C}$  such that

$$\left| \frac{f(\psi_{a,b}^{\text{DE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))} \right| \leq \tilde{C} |Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))| = \frac{\tilde{C} \{(b-a)/2\}^{2\alpha}}{\cosh^{2\alpha}(\pi \sinh(jh)/2)} \leq \tilde{C} (b-a)^{2\alpha} e^{-\pi \alpha \sinh(|jh|)}.$$

Furthermore using  $\sinh(|jh|) \geq (e^{|jh|} - 1)/2$ , and putting  $C_2 = \tilde{C} (b-a)^{2\alpha} e^{\frac{\pi}{2} \alpha}$ , we have

$$\begin{aligned} \sum_{|j| > N} \left| \frac{f(\psi_{a,b}^{\text{DE}}(jh))}{Q_{a,b}(\psi_{a,b}^{\text{DE}}(jh))} \right| &\leq C_2 \sum_{|j| > N} e^{-\frac{\pi}{2} \alpha \exp(|jh|)} \\ &= 2C_2 \sum_{j > N} e^{-\frac{\pi}{2} \alpha \exp(jh)} \\ &\leq 2C_2 \int_N^\infty e^{-\frac{\pi}{2} \alpha \exp(sh)} ds \\ &\leq 2C_2 \left\{ \frac{2}{\pi \alpha h e^{Nh}} \right\} \int_N^\infty \left\{ \frac{\pi \alpha h e^{sh}}{2} \right\} e^{-\frac{\pi}{2} \alpha \exp(sh)} ds \\ &= \frac{4C_2}{\pi \alpha h e^{Nh}} e^{-\frac{\pi}{2} \alpha \exp(Nh)}. \end{aligned} \quad (16)$$

Combining (15) with (16), we get (14). ■

Theorem 7 is then established by taking  $h$  as  $h = \log(2dN/\mu)/N$  in Lemma 6 and Lemma 7.

## 4 PROOFS OF THE THEOREMS IN SECTION 2

## 4.2.2 Proof of Theorem 6 (approximation error of integrals)

Theorem 6 can be shown in almost the same manner as the SE-Sinc case (Theorem 3).

Let us start with the next theorem.

**Theorem 9** (Tanaka et al. [20, Theorem 3.1]). Let  $(f/Q_{a,b}) \in \mathbf{L}_\beta(\psi_{a,b}^{\text{DE}}(\mathcal{D}_d))$  for  $d$  with  $0 < d < \pi/2$ . Let  $N$  be a positive integer with  $N > \beta/(4d)$ , and let  $h$  be selected by  $h = \log(4dN/\beta)/N$ . Then there exists a constant  $C$  independent of  $N$  such that

$$\left| \int_a^b F(s) ds - h \sum_{k=-N}^N F(\psi_{a,b}^{\text{DE}}(kh)) \{\psi_{a,b}^{\text{DE}}\}'(kh) \right| \leq C e^{-2\pi d N / \log(4dN/\beta)}.$$

Applying this theorem to the approximation  $\mathbf{I}_a^{1-p} g \approx \mathcal{I}_N^{\text{DE}} g$ , we have the next lemma.

**Lemma 8.** Assume that there exists a constant  $d$  with  $0 < d < \pi/2$  such that  $g$  is analytic and bounded uniformly on  $\psi_{a,t}^{\text{DE}}(\mathcal{D}_d)$  for all  $t \in [a, b]$ . Let  $\mu = \min\{1-p, \alpha\}$ ,  $N$  be a positive integer with  $N > \mu/(4d)$ , and  $h$  be selected by  $h = \log(4dN/\mu)/N$ . Then there exists a constant  $C$  independent of  $N$  such that

$$\max_{t \in [a, b]} \left| \mathbf{I}_a^{1-p}[g](t) - \mathcal{I}_N^{\text{SE}}[g](t) \right| \leq C e^{-2\pi d N / \log(4dN/\mu)}. \quad (17)$$

We can relax the condition on  $g$  using the following lemma.

**Lemma 9.** Let  $g$  be analytic and bounded on  $\psi_{a,b}^{\text{DE}}(\mathcal{D}_d) \cup \{b\}$  for  $d$  with  $0 < d < \pi/2$ . Then  $g$  is analytic and bounded uniformly on  $\psi_{a,t}^{\text{DE}}(\mathcal{D}_d)$  for all  $t \in [a, b]$ .

Since its proof is far more complicated than the SE-Sinc case (Lemma 2), we leave it to the end of this section. If we accept this lemma, Lemma 8 can be rewritten as follows.

**Lemma 10.** Let  $g$  be analytic and bounded on  $\psi_{a,b}^{\text{DE}}(\mathcal{D}_d) \cup \{b\}$  for  $d$  with  $0 < d < \pi/2$ . Let  $\mu = \min\{1-p, \alpha\}$ ,  $N$  be a positive integer with  $N > \mu/(4d)$ , and  $h$  be selected by  $h = \log(4dN/\mu)/N$ . Then there exists a constant  $C$  independent of  $N$  such that (17) holds.

Tomoaki Okayama et al.

If  $(f/Q_{a,b}) \in \mathbf{L}_\alpha(\psi_{a,b}^{\text{DE}}(\mathcal{D}_d))$  holds, then  $f'$  is analytic and bounded on  $\psi_{a,b}^{\text{DE}}(\mathcal{D}_{d-\epsilon})$  for any  $\epsilon$  with  $0 < \epsilon < d$ . Choosing  $\epsilon = d/2$  and using Lemma 10, we obtain Theorem 6.

It remains to prove Lemma 9. We commence by showing the following lemma. Here  $\zeta^*$  denotes a conjugate complex number of  $\zeta$ , and  $\text{sgn}$  denotes the so-called *sign function*, defined by

$$\text{sgn}(x) = \begin{cases} 1 & (x > 0), \\ 0 & (x = 0), \\ -1 & (x < 0). \end{cases}$$

**Lemma 11.** Let us define  $\mathcal{D}_d^+$  and  $\mathcal{D}_d^-$  as  $\mathcal{D}_d^+ = \{\zeta \in \overline{\mathcal{D}_d} : \text{Im } \zeta > 0\}$  and  $\mathcal{D}_d^- = \{\zeta \in \overline{\mathcal{D}_d} : \text{Im } \zeta < 0\}$ , where  $\mathcal{D}_d$  is defined by (10). Let  $f$  be a continuous function that satisfies  $f(\zeta^*) = \{f(\zeta)\}^*$  on  $\overline{\mathcal{D}_d}$ . Then the next two assertions are equivalent:

- (a) The function  $f$  satisfies that (a1)  $\text{Im}\{f(\zeta)\} = 0$  if and only if  $\text{Im } \zeta = 0$  on  $\overline{\mathcal{D}_d}$ , and (a2) there exists  $\zeta_0 \in \mathcal{D}_d^+$  such that  $\text{Im}\{f(\zeta_0)\} > 0$ .
- (b) The function  $f$  satisfies  $\text{sgn}[\text{Im}\{f(\zeta)\}] = \text{sgn}[\text{Im } \zeta]$  for all  $\zeta \in \overline{\mathcal{D}_d}$ , i.e.

$$\text{Im}\{f(\zeta)\} > 0 \quad \text{for all } \zeta \in \mathcal{D}_d^+, \quad (18)$$

$$\text{Im}\{f(\zeta)\} = 0 \quad \text{for all } \zeta \in \mathbb{R}, \quad (19)$$

$$\text{Im}\{f(\zeta)\} < 0 \quad \text{for all } \zeta \in \mathcal{D}_d^-. \quad (20)$$

*Proof.* We show only (a)  $\Rightarrow$  (b) since clearly (b)  $\Rightarrow$  (a) holds. The second condition (19) is obvious by the assumption (a1).

Suppose (18) does not hold, i.e. there exists  $\eta_0 \in \mathcal{D}_d^+$  such that  $\text{Im}\{f(\eta_0)\} \leq 0$ . Let  $\mathcal{C}$  be a closed arc defined by  $\mathcal{C} = \{\zeta = \lambda\zeta_0 + (1 - \lambda)\eta_0 : \lambda \in [0, 1]\}$ , where  $\zeta_0$  is given in the assumption (a2). Since  $\text{Im } f$  is continuous on  $\mathcal{C}$ , there exists  $\zeta \in \mathcal{C}$  such that  $\text{Im}\{f(\zeta)\} = 0$  by the intermediate-value theorem. This is contradictory to the assumption (a1). Thus the first condition (18) holds.



## 4 PROOFS OF THE THEOREMS IN SECTION 2

The third condition (20) can be shown in the same manner as (18), if we find some  $\xi_0 \in \mathcal{D}_d^-$  such that  $\operatorname{Im}\{f(\xi_0)\} < 0$ . By assumption it holds that

$$\operatorname{Im}\{f(\zeta_0)\} + \operatorname{Im}\{f(\zeta_0^*)\} = \operatorname{Im}[f(\zeta_0) + f(\zeta_0^*)] = \operatorname{Im}[f(\zeta_0) + \{f(\zeta_0)\}^*] = \operatorname{Im}[2 \operatorname{Re}\{f(\zeta_0)\}] = 0.$$

Thus  $\operatorname{Im}\{f(\zeta_0^*)\} < 0$ , which completes the proof.  $\blacksquare$

Let us define functions  $G_1$  and  $G_2$  as

$$G_1(\eta) = \eta + \sqrt{1 + \eta^2}, \quad G_2(\eta) = \frac{\sqrt{1 + \eta^2}}{1 + e^{-\pi\eta}}.$$

We can check that  $G_1$  satisfies the assumptions of Lemma 11 and the assertion (a). Hence we have the next lemma from the assertion (b).

**Lemma 12.** The equality  $\operatorname{sgn}[\operatorname{Im}\{G_1(\eta)\}] = \operatorname{sgn}[\operatorname{Im} \eta]$  holds for all  $\eta \in \overline{\mathcal{D}_1}$ .

In fact it holds for all  $\eta \in \mathbb{C}$ , but  $\eta \in \overline{\mathcal{D}_1}$  is sufficient for our purpose.

Lemma 11 can not be applied directly to the function  $G_2$  since  $G_2(\pm i) = \pm\infty$ , i.e.  $G_2$  is not continuous at  $\eta = \pm i$ . However,  $G_2$  still satisfies  $\operatorname{Im}\{G_2(+i)\} = +\infty > 0$  and  $\operatorname{Im}\{G_2(-i)\} = -\infty < 0$ . Therefore  $\operatorname{sgn}[\operatorname{Im}\{G_2(\eta)\}] = \operatorname{sgn}[\operatorname{Im}\eta]$  holds even if  $\eta = \pm i$ . Then we have the next lemma.

**Lemma 13.** The equality  $\operatorname{sgn}[\operatorname{Im}\{G_2(\eta)\}] = \operatorname{sgn}[\operatorname{Im} \eta]$  holds for all  $\eta \in \overline{\mathcal{D}_1}$ .

Using the two lemmas above, we prove Lemma 9.

*Proof.* In the same argument as in Lemma 2, we consider the function  $\cos(\operatorname{Im} \zeta(t))$ , where  $\zeta(t) = \{\psi_{a,t}^{\text{DE}}\}^{-1}(z)$ , and show it is a monotonically increasing function. Since

$$\frac{d}{dt} \cos(\operatorname{Im}(\zeta(t))) = -\sin(\operatorname{Im} \zeta(t)) \frac{d}{dt} \{\operatorname{Im} \zeta(t)\},$$

we examine the signs of  $\sin(\operatorname{Im} \zeta(t))$  and  $\{\operatorname{Im} \zeta\}'$  below. Let us define a function  $\eta(t)$  as

Tomoaki Okayama et al.

$\eta(t) = \frac{1}{\pi} \log \left( \frac{z-a}{t-z} \right)$ . Then  $\zeta(t) = \log\{\eta(t) + \sqrt{1 + \eta^2(t)}\}$ , and

$$\sin(\operatorname{Im} \zeta(t)) = \frac{\operatorname{Im} [\eta(t) + \sqrt{1 + \eta^2(t)}]}{|\eta(t) + \sqrt{1 + \eta^2(t)}|},$$

because  $\sin(\arg(\xi)) = \operatorname{Im} \xi / |\xi|$  for all  $\xi \in \mathbb{C}$ . We set  $x_t = \operatorname{Re} \eta(t)$  and  $y_t = \operatorname{Im} \eta(t)$  here.

Note  $\eta(t) \in \overline{\mathcal{D}_1}$  by definition. According to Lemma 12, it follows for all  $\eta(t) \in \overline{\mathcal{D}_1}$  that

$$\operatorname{sgn} \left\{ \operatorname{Im} \left[ \eta(t) + \sqrt{1 + \eta^2(t)} \right] \right\} = \operatorname{sgn} \{ \operatorname{Im} [G_1(\eta(t))] \} = \operatorname{sgn} \{y_t\}. \quad (21)$$

Next we examine the sign of  $\{\operatorname{Im} \zeta\}'$ . The function  $\operatorname{Im} \zeta(t)$  can be written as

$$\operatorname{Im} \zeta(t) = \frac{1}{2i} \left[ \log \left\{ \eta(t) + \sqrt{1 + \eta^2(t)} \right\} - \log \left\{ \eta^*(t) + \sqrt{1 + \{\eta^*(t)\}^2} \right\} \right].$$

By differentiating and rewriting this equation, we obtain

$$\frac{d}{dt} \operatorname{Im} \zeta(t) = \frac{\operatorname{Im} [(t-z) \sqrt{1 + \eta^2(t)}]}{|(t-z) \sqrt{1 + \eta^2(t)}|^2}.$$

From the definition of  $\eta(t)$ , we have  $(t-z) = (t-a)/(1 + e^{\pi\eta(t)})$ , and then

$$\frac{d}{dt} \operatorname{Im} \zeta(t) = \left| \frac{1 + e^{\pi\eta(t)}}{\sqrt{(t-a)\{1 + \eta^2(t)\}}} \right|^2 \operatorname{Im} \left[ \frac{\sqrt{1 + \eta^2(t)}}{1 + e^{\pi\eta(t)}} \right].$$

Thus by applying Lemma 13, it follows for all  $\eta(t) \in \overline{\mathcal{D}_1}$  that

$$\operatorname{sgn} \left\{ \operatorname{Im} \left[ \frac{\sqrt{1 + \eta^2(t)}}{1 + e^{\pi\eta(t)}} \right] \right\} = \operatorname{sgn} \{ \operatorname{Im} [G_2(-\eta(t))] \} = -\operatorname{sgn} \{y_t\}. \quad (22)$$

## REFERENCES

Finally, using the expressions (21) and (22), we obtain the desired conclusion:

$$\operatorname{sgn} \left\{ \frac{d}{dt} \cos(\operatorname{Im}(\zeta(t))) \right\} = \frac{\{\operatorname{sgn}(y_t)\}^2}{\operatorname{sgn} \left\{ \left| \eta(t) + \sqrt{1 + \eta^2(t)} \right| \right\}} \operatorname{sgn} \left\{ \left| \frac{1 + e^{\pi \eta(t)}}{\sqrt{(t-a)\{1 + \eta^2(t)\}}} \right|^2 \right\} \geq 0. \quad \blacksquare$$

## Acknowledgement

This study was supported by Global COE Program “The research and training center for new development in mathematics,” MEXT, Japan.

## References

- [1] G. A. ANASTASSIOU, Opial type inequalities involving Riemann–Liouville fractional derivatives of two functions with applications, *Math. Comput. Modelling*, 48, 344–374 (2008).
- [2] A. CARPINTERI, F. MAINARDI (eds.), *Fractals and Fractional Calculus in Continuum Mechanics*, Springer-Verlag, Wien, 1997.
- [3] S. DAS, *Functional Fractional Calculus for Systems Identification and Controls*, Springer, Berlin, 2007.
- [4] K. DIETHELM, An investigation of some nonclassical methods for the numerical approximation of Caputo-type fractional derivatives, *Numer. Algorithms*, 47, 361–390 (2008).
- [5] K. DIETHELM, N. J. FORD, A. D. FREED, Y. LUCHKO, Algorithms for the fractional calculus: A selection of numerical methods, *Comput. Methods Appl. Mech. Engrg.*, 194, 743–773 (2005).
- [6] R. HILFER (ed.), *Applications of Fractional Calculus in Physics*, World Scientific, Singapore, 2000.
- [7] A. A. KILBAS, H. M. SRIVASTAVA, J. J. TRUJILLO, *Theory and Applications of Fractional Differential Equations*, Elsevier, Amsterdam, 2006.

Tomoaki Okayama et al.

- [8] CH. LUBICH, Runge–Kutta theory for Volterra and Abel integral equations of the second kind, *Math. Comp.*, 41, 87–102 (1983).
- [9] R. L. MAGIN, *Fractional Calculus in Bioengineering*, Begell House, Connecticut, 2006.
- [10] M. MORI, A. NURMUHAMMAD, T. MURAI, Numerical solution of Volterra integral equations with weakly singular kernel based on the DE-sinc method, *Japan J. Indust. Appl. Math.*, 25, 165–183 (2008).
- [11] M. MORI, M. SUGIHARA, The double-exponential transformation in numerical analysis, *J. Comput. Appl. Math.*, 127, 287–296 (2001).
- [12] T. OKAYAMA, T. MATSUO, M. SUGIHARA, Error estimates with explicit constants for Sinc approximation, Sinc quadrature and Sinc indefinite integration, Mathematical Engineering Technical Reports, 2009-01, The University of Tokyo, 2009.
- [13] T. OKAYAMA, T. MATSUO, M. SUGIHARA, Sinc-collocation methods for weakly singular Fredholm integral equations of the second kind, Mathematical Engineering Technical Reports, 2009-02, The University of Tokyo, 2009.
- [14] I. PODLUBNY, *Fractional Differential Equations*, Academic Press, San Diego, 1999.
- [15] B. V. RILEY, The numerical solution of Volterra integral equations with nonsmooth solutions based on sinc approximation, *Appl. Numer. Math.*, 9, 249–257 (1992).
- [16] J. SABATIER, O. P. AGRAWAL, J. A. T. MACHADO (eds.), *Advances in Fractional Calculus: Theoretical Developments and Applications in Physics and Engineering*, Springer, Dordrecht, 2007.
- [17] F. STENGER, *Numerical Methods Based on Sinc and Analytic Functions*, Springer-Verlag, New York, 1993.
- [18] M. SUGIHARA, T. MATSUO, Recent developments of the Sinc numerical methods, *J. Comput. Appl. Math.*, 164/165, 673–689 (2004).
- [19] H. SUGIURA, T. HASEGAWA, Quadrature rule for Abel’s equations: Uniformly approximating fractional derivatives, *J. Comput. Appl. Math.*, 223, 459–468 (2009).
- [20] K. TANAKA, M. SUGIHARA, K. MUROTA, M. MORI, Function classes for double exponential integration formulas, *Numer. Math.*, 111, 631–655 (2009).

# Boundary type quadrature formulas over axially symmetric regions

Tian-Xiao He

Dept. Math & CS, Illinois Wesleyan University  
Bloomington, IL 61702-2900

## Abstract

A boundary type quadrature formula (BTQF) is an approximate integration formula with all its evaluation points lying on the boundary of the integration domain. This type formulas are particularly useful for the cases when the values of the integrand functions and their derivatives inside the domain are not given or are not easily determined. In this paper, we will establish the BTQFs over some axially symmetric regions. We will discuss the following three questions in the construction of BTQFs: (i) What is the highest possible degree of algebraic precision of the BTQF if it exists? (ii) What is the fewest number of the evaluation points needed to construct a BTQF with the highest possible degree of algebraic precision? (iii) How to construct the BTQF with the fewest evaluation points and the highest possible degree of algebraic precision?

## 1 Introduction

Although numerical multivariate integration is an old subject, it has never been applied as widely as it is now. We can find its applications everywhere in math, science, and economics. A good example might be the collateralized mortgage obligation (CMO), which can be formulated as a multivariate integral over the 180-dimensional unit cube ([2]). A boundary quadrature formula is an approximate integration formula with all its evaluation points lying on the boundary of the domain of integration. Such a formula may be particularly useful for the cases when the values of the integrand function and its derivatives inside the domain are not given or are not easily determined.

Indeed, boundary quadrature formulas are not really new. From the viewpoint of numerical analysis, the classical Euler-Maclaurin summation formula and the Hermite two-end multiple nodes quadrature formulas may be regarded as one-dimensional boundary quadrature formulas since they make use of only the integrand function values and their derivatives at the limits of integration. The earliest example of a boundary quadrature formula with some algebraic precision for multivariate integration is possibly the formula of algebraic precision (or degree) 5 for a triple integral over a cube given by Sadowsky [30] in 1940. He used 42 points

on the surface of a cube to construct the quadrature, which has been modified by the author with a quadrature of 32 points, the fewest possible boundary points (see [9] and [10]). Some 20 years later after Sadowsky's work, Levin [26] and [27], Federenko [6], and Ionescu [21] investigated individually certain optimal boundary quadrature formulas for double integration over a square using partial derivatives at some boundary points of the region. Despite these advances, however, both the general principle and the general technique for construction remained lacking for many years.

During 1978-87, based on the ideas of the dimension-reducing expansions (DRE) of multivariate integration shown in Hsu 1962 and 1963, Hsu, Wang, Zhou, Yang, and the author developed a general process for the construction of BTQFs in [17]-[20] and [9]-[15].

The analytic approach for constructing BTQFs is based on the dimension-reducing expansions (DRE), which reduces a higher dimensional integral to lower dimensional integrals with or without a remainder. Hence, a type of boundary quadratures can be constructed by using the expansions.

The DRE without remainder is also called an exact DRE. Obviously, a DRE can be used to reduce the computation load of many very high dimensional numerical integration's, such as the CMO problem mentioned above. Most DRE's are based on Green's Theorem in real or complex field. In 1963, using the theorem, Hsu [17] devised a way to construct a DRE with algebraic precision (degree of accuracy) for multivariate integrations. From 1978 to 1986, Hsu, Zhou, and the author (see [18], [19], [20], and [?]) developed a more general method to construct a DRE with algebraic precision and estimate its remainder. In 1972, with the aid of Green's Theorem and the Schwarz function, P.J. Davis [4] gave an exact DRE for a double integral over a complex field. In 1979, also by using Green's Theorem, Kratz [24] constructed an exact DRE for a function that satisfied a type of partial differential equations. Lastly, if we want this introduction to be complete, we must not overlook Burrows' DRE for measurable functions. His DRE can reduce a multivariate integration into an one dimensional integral. Some important applications of DRE include the construction of BTQFs and asymptotic formulas for oscillatory integrals, for instance, the integrals on spheres,  $S^d = \{x \in R^d : |x| = 1\}$  and balls,  $B^d = \{x \in R^d : |x| \leq 1\}$ , presented by Kalnins, Miller, Jr., and Tratnik [22], Lebedev and Skorokhodov [25], Mhaskar, Narcowich, and Ward [28], Xu [35], etc.

In this paper, we will discuss the algebraic approach to constructing BTQFs for a multiple integral over a bounded closed region  $\Omega$  in  $\mathbb{R}^n$ , which is of the form

$$\int_{\Omega} w(X)f(X)dX.$$

In this expression,  $w(X)$  and  $f(X)$  are continuous on  $\Omega$ , and  $w(X)$  is the weight function. ( $w(X)$  can be 1 particularly.) We are seeking the BTQF of the integral with the form

$$\int_{\Omega} w(X)f(X)dX \approx \sum_{0 \leq m_1 + \dots + m_n \leq m} \sum_{i \in I} a_i^{m_1, \dots, m_n} D^{m_1, \dots, m_n} f(X_i), \quad (1)$$

where  $dX$  is the volume measure;  $a_i^{m_1, \dots, m_n}$  ( $i \in I$  and  $0 \leq m_1 + \dots + m_n \leq m$ ) are real or complex quadrature coefficients;  $D^{m_1, \dots, m_n} = \partial^{m_1 + \dots + m_n} / \partial x_1^{m_1} \dots \partial x_n^{m_n}$ ;

and  $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n})$  ( $i \in I$ ) are evaluation points (or nodes) of  $f$  on  $\partial\Omega$ , the boundary of  $\Omega$ . In particular, when  $m = 0$  we write  $a_i^{m_1, \dots, m_n} = a_i$  and formula (1) can be rewritten as

$$\int_{\Omega} w(X) f(X) dX \approx \sum_{i \in I} a_i f(X_i). \quad (2)$$

(2) is called a *BTQF without derivative terms*. When  $m \neq 0$ , (1) is called a *BTQF with derivative terms*. The corresponding error functionals of approximations (1) and (2) are defined respectively by

$$E(f) \equiv E(f; \Omega) = \int_{\Omega} w(X) f(X) dX - \sum_{0 \leq m_1 + \dots + m_n \leq m} \sum_{i \in I} a_i^{m_1, \dots, m_n} D^{m_1, \dots, m_n} f(X_i) \quad (3)$$

and

$$E(f) \equiv E(f; \Omega) = \int_{\Omega} w(X) f(X) dX - \sum_{i \in I} a_i f(X_i). \quad (4)$$

Suppose that  $\partial\Omega$  can be described by a system of parametric equations. In particular, the points  $X = (x_1, \dots, x_n)$  on  $\partial\Omega$  satisfy the equation

$$\Phi(X) = 0, \quad (5)$$

where  $\Phi$  has continuous partial derivatives. In addition,  $\Phi(X) \leq 0$  for all points in  $\Omega$ .

Let  $S$  be another region in  $\mathbb{R}^n$ , and let  $J : Y = JX$ ,  $X \in \Omega$ , be a transform from  $\Omega$  to  $S$  with positive Jacobian

$$|J| = \left| \frac{\partial(Y)}{\partial(X)} \right| > 0,$$

$X \in \Omega$ .  $J$  is one-to-one and has the inverse  $J^{-1} : X = J^{-1}Y$ ,  $Y \in S$ . Denote  $w_1(Y) = w_1(JX) = w(X)$ . Then for any continuous function  $g(X)$

$$\int_S w_1(Y) g(Y) dY = \int_{\Omega} w_1(Y) g(Y) |J| dX.$$

Denoting  $Y_i = JX_i$  ( $i \in I$ ),  $|J_i| = |J|_{X=X_i}$ , and taking  $f(X) = |J|g(Y) = |J|g(JX)$  in equation (4), we obtain

$$E(|J|g; \Omega) = \int_{\Omega} w(X) |J| g(Y) dX - \sum_{i \in I} a_i |J_i| g(Y_i) = \int_S w_1(Y) g(Y) dY - \sum_{i \in I} b_i g(Y_i),$$

where  $b_i = a_i |J_i|$  ( $i \in I$ ). Obviously, if  $Y$ , the boundary points of  $S$ , satisfy  $\Phi_1(Y) = \Phi_1(JX) = \Phi(X) = 0$ , then  $J$  maps the boundary evaluation points  $X_i$  ( $i \in I$ ) on  $\Omega$  onto the boundary evaluation points  $Y_i = JX_i$  on  $S$ . Consequently, we have the following result.

**Theorem 1** *Let the error functional of the quadrature formula*

$$\int_S w_1(Y)g(Y)dY \approx \sum_{i \in I} b_i g(Y_i) \quad (6)$$

*be  $E(g; S) = \int_S w_1(Y)g(Y)dY - \sum_{i \in I} b_i g(Y_i)$ . Then  $E(g; S) = E(|J|g; \Omega)$ . In particular, if  $|J|$  is a constant, then  $E(g; S) = |J|E(g; \Omega)$ . In this case,  $E(g; \Omega) = 0$  implies  $E(g; S) = 0$ .*

*In addition, if the boundary of  $S$  is defined by  $\Phi_1(Y) = \Phi_1(JX) = \Phi(X) = 0$  and  $\Phi(X) = 0$  defines the boundary of  $\Omega$ , then quadrature formula (6) is also a BTQF.*

In this paper, we will establish the BTQFs over some axially symmetric regions or fully symmetric regions (see the definitions below). Theorem 1 tells us that we can construct the BTQFs over many more regions from the obtained BTQFs over the special regions by using certain transforms. In addition, if the transform is linear, then the new BTQF is of the same algebraic precision degree as the old BTQF.

## 2 BTQFs without derivatives

Three questions arise during the construction of BTQFs (1):

- (i) What is the highest possible degree of algebraic precision of the BTQF if it exists?
- (ii) What is the fewest number of the evaluation points needed to construct a BTQF with the highest possible degree of algebraic precision?
- (iii) How to construct the BTQF with the fewest evaluation points with the fewest evaluation points and the highest possible degree of algebraic precision?

We now answer the first question. In most cases, BTQF (1) has an inherent highest degree of algebraic precision. For instance, if  $\Phi(X)$  is a polynomial of degree  $m$ , then the highest possible degree of algebraic precision of the BTQF without derivative terms (i.e., formula (2)) cannot exceed  $m - 1$  because the summation on the right-hand side of (2) becomes zero and the integral value on the left-hand side is negative when  $f = \Phi$ . Hence, when the boundary function  $\Phi$  is a polynomial of a low degree, to raise the degrees of algebraic precision of the quadrature formulas, we must construct BTQFs with derivative terms (i.e., formula (1) with  $m \neq 0$ ).

In the following, we are going to find the solutions to questions (ii) and (iii). To simplify our discussion, we limit the region in question,  $\Omega$ , to be axially symmetric or fully symmetric. An *axially symmetric region* is a region that for any point  $X = (x_1, \dots, x_n)$  in it, must contain all points with the form  $(\pm x_1, \dots, \pm x_n)$ . The set of axially symmetric points associated with  $X$  forms a reflection group. If a region containing a point  $X = (x_1, \dots, x_n)$  also contains all points  $(\pm a_1, \dots, \pm a_n)$ , where  $(a_1, \dots, a_n)$  is a permutation of  $(x_1, \dots, x_n)$ , then the region is called a *fully symmetric region*. Throughout, we will denote all fully symmetric points,  $(\pm a_1, \dots, \pm a_n)$ , associated with  $X$  by  $X_{FS}$  and call  $X$  the generator of the fully symmetric point set. The cardinal number of the set of fully symmetric points



associated with a generator  $X \in \mathbb{R}^n$  is  $2^n(n!)$ . Obviously, a fully symmetric region is an axially symmetric region, but the converse is not true.

A quadrature formula is called a *fully symmetric quadrature formula* if the quadrature sum can be divided into several subsums such that in each of the subsums, the evaluation points are fully symmetric and the corresponding quadrature coefficients are the same. In addition, if the fully symmetric evaluation points are on the boundary of the integral region, then the corresponding quadrature formula is called a *fully symmetric BTQF*.

Denote a monomial in terms of  $X$  by  $X^\alpha$  ( $\alpha \in \mathbb{Z}_0^n$ ), which can be written in the form  $X^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ , where  $(\alpha_1, \dots, \alpha_n)$  is called the exponent of  $X^\alpha$ .

From the definition of the fully symmetric region, we immediately have the following results.

**Theorem 2** *The value of a multiple integral of a monomial  $X^\alpha$  over an axially symmetric region is zero if  $\alpha$  contains an odd component. The value of a multiple integral of  $X^\alpha$  over a fully symmetric region depends on  $\alpha$ , but is independent of the order of  $\alpha_i$  ( $i = 1, \dots, n$ ).*

**Theorem 3** *Denote by  $\pi_r^n(X)$  the set of all polynomials of degree no greater than  $r$ . Let  $\Omega$  be a fully symmetric region,*

$$\int_{\Omega} f(X) dX \approx \sum_{i \in I} a_i f(X_i) \quad (7)$$

*be a fully symmetric BTQF, and  $E : f \rightarrow \mathbb{R}$  be the error operator defined by*

$$E(f) \equiv E(f; \Omega) = \int_{\Omega} f(X) dX - \sum_{i \in I} a_i f(X_i).$$

*(The above expression is a special form of (4) with  $w(X) = 1$ .) Then  $\pi_{2k+1}^n \subset N(E)$ , the null space of  $E$ , if and only if*

$$x_1^{2k_1} \cdots x_n^{2k_n} \in N(E) \quad 0 \leq k_1 \leq \cdots \leq k_n, \quad k_1 + \cdots + k_n \leq k. \quad (8)$$

Theorem 3 can be considered as the general principle for constructing fully symmetric BTQFs. First, we set one or more sets of fully symmetric evaluation points, with possibly some unknown points  $\{X_i\}$ , on the boundary  $\partial\Omega$  and assume the quadrature coefficients  $a_i$  corresponding to each set to be the same. Then substituting all  $f(X) = x_1^{2k_1} \cdots x_n^{2k_n}$  ( $0 \leq k_1 \leq \cdots \leq k_n$  and  $k_1 + \cdots + k_n \leq k$ ) into  $E(f) = 0$ , we obtain a system about  $X_i$  and  $a_i$ . Finally, we solve the system for  $X_i$  and  $a_i$  and a quadrature formula is constructed. However, a fully symmetric quadrature formula usually has too many evaluation points. (Remember that for a point  $X \in \mathbb{R}^n$  there are, in general,  $2^n(n!)$  fully symmetric points.) In order to reduce the number of evaluation points in the quadrature formula, we can use an alternative form of Theorem 3 to construct a different type of symmetric quadrature formulas. We will use the following example to illustrate the idea.

**Example 1.** Consider a triple integral over the region  $C_3 = [-1, 1]^3$ . Obviously, the inherent highest degree of algebraic precision of the BTQF

is 5. To construct a fully symmetric BTQF, we make use of the following fully symmetric evaluation points.

$$(1, 0, 0)_{FS}, \quad (1, 1, 0)_{FS}, \quad \text{and} (1, x_0, x_0)_{FS},$$

where  $x_0$  ( $0 < x_0 < 1$ ) is undetermined. The three sets of fully symmetric points contain a total of 42 points (6, 12, and 24 points for the first, second, and third set respectively). Let the respective quadrature coefficients for each set of fully symmetric points be  $L$ ,  $M$ , and  $N$ , all of which can be found using the general principle for constructing fully symmetric BTQFs. Substitute  $f = 1$ ,  $x^2$ ,  $x^4$ , and  $x^2y^2$  into

$$\int_{C_3} f(x, y, z) dx dy dz = a_1 \sum f_6 + a_2 \sum f_{12} + a_3 \sum f_{24},$$

where  $\sum f_6$ ,  $\sum f_{12}$ , and  $\sum f_{24}$  are the sums of the function values of  $f$  over the first, second, and third set of symmetric points, respectively. Solving the above system yields

$$x_0 = \sqrt{\frac{5}{8}}, \quad a_1 = \frac{364}{225}, \quad a_2 = -\frac{160}{225}, \quad a_3 = \frac{64}{225},$$

giving the following BTQF of algebraic precision order 5.

$$\int_{C_3} f(x, y, z) dx dy dz \approx \frac{4}{225} \left[ 91 \sum f_6 - 40 \sum f_{12} + 16 \sum f_{24} \right]. \quad (9)$$

Quadrature formula (9), given by Sadowsky [30], uses too many evaluation points. Carefully considering Theorem 3, we find that the principle of constructing fully symmetric BTQFs shown in the theorem can be used to construct some “partial” symmetric BTQFs with fewer evaluation points.

A set of points  $X_i \in \mathbb{R}^n$  ( $i \in I$ ) is called a *symmetric point set of degree  $k$*  if it possesses the following two properties.

(a)  $\sum_{i \in I} f(X_i) = 0$  for all  $f(X) = X^\alpha$ , where  $\alpha$  contains an odd component.

(b)  $\sum_{i \in I} f(X_i)$  are the same for all  $f(X) = x_1^{2k_1} \cdots x_n^{2k_n}$ ,  $2(k_1 + \cdots + k_n) = r$ . Here,  $r \leq k$ .

Obviously, a set of fully symmetric points must be a set of symmetric points of any degree, but the converse is not true. For instance, a symmetric point set of degree 5 may not be a fully symmetric point set. We now list all symmetric point sets of degree 5 on the boundary of  $C_3$  as follows.  $I = \{(\pm 1, \pm x_0, 0), (\pm x_0, 0, \pm 1), (0, \pm 1, \pm x_0), 0 < x_0 < 1\}$ ,  $II = \{(\pm y_0, \pm 1, 0), (\pm 1, 0, \pm y_0), (0, \pm y_0, \pm 1), 0 < y_0 < 1\}$ ,  $III = \{(\pm 1, \pm 1, 0), (\pm 1, 0, \pm 1), (0, \pm 1, \pm 1)\}$ ,  $IV = \{(\pm 1, \pm 1, \pm 1)\}$ ,  $V = \{(1, 0, 0)_{FS}\}$ ,  $VI = \{(1, x_1, x_2)_{FS}, 0 < x_1, x_2 < 1\}$ ,  $VII = \{(1, 1, x_3)_{FS}, 0 < x_3 < 1\}$ , where sets  $V$ ,  $VI$ , and  $VII$  are fully symmetric, but others are not.

If a BTQF constructed by using symmetric point set of degree  $k$  satisfies condition (8), then it is called a *symmetric BTQF of degree  $k$* .

**Example 2.** As an example, we now use the sets  $I$ ,  $III$ , and  $IV$  to construct a symmetric BTQF of degree 5 with 32 evaluation points over  $C_3$ . Denote the quadrature coefficients corresponding to  $I$ ,  $III$ , and  $IV$  as  $a_1$ ,  $a_2$ , and  $a_3$  respectively.

Following the procedure shown in Example 1, we obtain a symmetric BTQF of degree 5 as follows

$$\begin{aligned} & \int_{C_3} f(x, y, z) dx dy dz \\ & \approx \frac{1}{63} \left[ 80 \sum f_{12}(I) - 52 \sum f_{12}(III) + 21 \sum f_8(IV) \right], \end{aligned} \quad (10)$$

$\sum f_{12}(I)$ ,  $\sum f_{12}(III)$ , and  $\sum f_8(IV)$  are the sums of the function values of  $f$  over the symmetric point sets  $I$ ,  $III$ , and  $IV$ , respectively; the numbers in the sub-indices are the cardinal numbers of the corresponding set.

Similarly, we can use sets  $II$ ,  $III$ , and  $IV$  to construct another symmetric BTQF of degree 5.

$$\begin{aligned} & \int_{C_3} f(x, y, z) dx dy dz \\ & \approx \frac{1}{63} \left[ 80 \sum f_{12}(II) - 52 \sum f_{12}(III) + 21 \sum f_8(IV) \right], \end{aligned} \quad (11)$$

where  $y_0 = \sqrt{\frac{3}{10}}$  in set  $II$ .

Quadratures (10) and (11) can be considered as two special cases of the following symmetric BTQF of degree 5, which is constructed by using  $I$ ,  $II$ , and  $IV$ .

$$\begin{aligned} \int_{C_3} f(x, y, z) dx dy dz & \approx \frac{4(1 + y_0^2)}{9(y_0^2 - x_0^2)} \sum f_{12}(I) \\ & + \frac{4(1 + x_0^2)}{9(x_0^2 - y_0^2)} \sum f_{12}(II) + \frac{1}{3} \sum f_8(IV), \end{aligned} \quad (12)$$

where

$$\sqrt{\frac{3}{10}} \leq y \leq 1, \quad y_0 \neq \sqrt{\sqrt{\frac{13}{5}} - 1}, \quad \text{and} \quad x_0 = \sqrt{\frac{8 - 5y_0^2}{5(1 + y_0^2)}}.$$

When  $y_0 = 1$  and  $y_0 = \sqrt{\frac{3}{10}}$  we obtain formulas (10) and (11), respectively.

It can be proved that the minimum number of evaluation points of symmetric BTQFs is 32. Since the quadrature formula is symmetric, on each boundary plane we must have the same number of evaluation points. Let the number of evaluation points on each boundary plane be  $k = 2$  (Obviously,  $k$  cannot be 1). The symmetric point set has to be  $I$  or  $II$ . It is easy to check that the sets cannot yield a symmetric BTQF of degree 5. Similarly, for the cases of  $k = 3, \dots, 9$ , no matter which symmetric point sets are chosen from  $\{I, \dots, VII\}$ , we find that there does not exist any symmetric BTQFs of degree 5 with evaluation points less than 32. For  $k \geq 10$ , every symmetric BTQF of degree 5, if it exists, must have more than 32 evaluation points. Thus, we obtain the following proposition.

**Proposition 4** *There exist infinitely many symmetric BTQFs of degree 5 with 32 evaluation points. In addition, the number of evaluation points of a symmetric BTQFs of degree 5 can not be less than 32.*

For BTQFS of degree 3, the minimum number of the evaluation points is reduced to 6. As an example, we give the following formula.

$$\int_{C_3} f(x, y, z) dx dy dz \approx \frac{4}{3} [f(1, 0, 0) + f(-1, 0, 0) + f(0, 1, 0) + f(0, -1, 0) + f(0, 0, 1) + f(0, 0, -1)].$$

**Example 3.** We will use a double layered spherical shell as an example to demonstrate the techniques of regrouping evaluation points to obtain the symmetric BTQF with the fewest evaluation points. A double layered spherical shell in  $\mathbb{R}^n$ , denoted by  $Sh_n$ , is defined by

$$Sh_n = \{X \in \mathbb{R}^n : a^2 \leq |X| \leq b^2\}.$$

It is easy to find that the largest degree of algebraic precision of BTQFs over  $Sh_n$  without derivatives is 3. We choose the following point sets as evaluation points:  $VIII = \{(\pm b, 0, \dots, 0), (0, \pm b, 0, \dots, 0), \dots, (0, \dots, 0, \pm b, 0)\}$ ,  $IX = \{(0, \dots, 0, \pm b)\}$ ,  $X = \{(0, \dots, 0, \pm a)\}$ .

Obviously, these sets are neither fully symmetric point sets nor symmetric point sets of degree 3, but by using these sets, we can construct a BTQF of degree 3 over  $Sh_n$  with the fewest evaluation points. Denote the quadrature coefficients corresponding to  $VIII$ ,  $IX$ , and  $X$  by  $a_1$ ,  $a_2$ , and  $a_3$ , respectively. The BTQF generated,

$$\int_{Sh_n} f(X) dX \approx a_1 \sum f_{2(n-1)}(VIII) + a_2 \sum f_2(IX) + a_3 \sum f_2(X), \quad (13)$$

is of algebraic precision of degree 3 if it holds exactly for  $f = 1$ ,  $x_1^2$ , and  $x_n^2$ ; i.e., coefficients  $a_i$  ( $i = 1, 2, 3$ ) have to be

$$\begin{aligned} a_1 &= \alpha(b^2 - a^2)(b^{n+2} - a^{n+2}) \\ a_2 &= \alpha(b^{n+4} + (n+1)a^{n+2}b^2 - 3b^{n+2}a^2 - (n-1)a^{n+4}) \\ a_3 &= \alpha b^2(2b^{n+2} - (n+2)a^n b^2 + na^{n+2}), \end{aligned}$$

where

$$\alpha = \frac{\pi^{n/2}}{2b^2\Gamma(\frac{n}{2} + 1)(n+2)(b^2 - a^2)}.$$

When  $n = 2$  and 3, formula (13) gives BTQFs over a ring domain and a 3-dimensional double layered spherical shell respectively as follows.

$$\begin{aligned} & \int_{Sh_2} f(x, y) dx dy \\ & \approx \frac{\pi(b^2 - a^2)}{8b^2} \{ (b^2 + a^2)[f(b, 0) + f(-b, 0)] + 2b^2[f(0, a) + f(0, -a)] \\ & \quad + (b^2 - a^2)[f(0, b) + f(0, -b)] \} \end{aligned}$$

$$\begin{aligned}
& \int_{Sh_3} f(x, y, z) dx dy dz \\
& \approx \frac{2\pi}{15b^2(b^2 - a^2)} \{ (b^2 - a^2)(b^5 - a^5)[f(b, 0, 0) + f(-b, 0, 0) + f(0, b, 0) \\
& \quad + f(0, -b, 0)] + b^2(2b^5 - 5a^3b^2 + 3a^5)[f(0, 0, a) + f(0, 0, -a)] \\
& \quad + (b^7 - 3a^2b^5 + 4a^5b^2 - 2a^7)[f(0, 0, b) + f(0, 0, -b)] \}.
\end{aligned}$$

Taking the limit  $a \rightarrow 0$ , from quadrature formula (13) we obtain the following quadrature formula over the sphere  $S_3$ , which has the algebraic precision of degree 3.

$$\begin{aligned}
& \int_{S_3} f(x, y, z) dx dy dz \approx \frac{\pi^{n/2} b^n}{2(n+2)\Gamma(\frac{n}{2} + 1)} \\
& \times \left( \sum f_{2(n-1)}(VIII) + \sum f_2(IX) + 4f(0, \dots, 0) \right).
\end{aligned}$$

We now prove that BTQF (13) is a formula with the fewest evaluation points.

**Theorem 5** *The minimum number of evaluation points of BTQFs over an  $n$ -dimensional double layered spherical shell  $Sh_n$  is  $2(n+1)$ . In particular, the minimum number of evaluation points for BTQFs over a ring domain and a 3-dimensional double layered spherical shell are respectively 6 and 8.*

*Proof.* For a BTQF over  $Sh_n$  with precision degree 3, we will first prove that the minimum number of evaluation points on the outside layer of  $Sh_n$  cannot be less than  $2n$ . Without a loss of generality, we assume that the number of evaluation points on the outside layer is  $2n-1$ . (The cases when the minimums are less than  $2n-1$  can be proved similarly.) We will see that a contradiction from this assumption. If the assumption is valid, we take the limit  $a \rightarrow 0$  to the BTQF and obtain a quadrature formula over an  $n$ -dimensional sphere with  $2n-1$  evaluation points as follows.

$$\int_{S_n} f(X) dX \approx a_0 f(0, \dots, 0) + \sum_{i=1}^{2n-1} a_i f(X_i), \quad (14)$$

where  $X_i$  ( $i = 1, \dots, 2n-1$ ) lie on the sphere surface and  $a_i \neq 0$  ( $i = 1, \dots, n$ ). We will prove it cannot be of algebraic precision degree 3.

Let us consider the following  $2n$  complex vectors

$$AX_1, \dots, AX_n, Ax_1^2, AX_2^2, \dots, AX_n^2, \quad (15)$$

where

$$AX_i = (\sqrt{a_1}x_{1,i}, \sqrt{a_2}x_{2,i}, \dots, \sqrt{a_{2n-1}}x_{2n-1,i})$$

and

$$AX_i^2 = (\sqrt{a_1}x_{1,i}^2, \sqrt{a_2}x_{2,i}^2, \dots, \sqrt{a_{2n-1}}x_{2n-1,i}^2).$$

Assume that there exist constants  $b_i$  ( $i = 1, \dots, 2n$ ) such that

$$\begin{aligned}
& b_1 AX_1 + \dots + b_n AX_n \\
& + b_{n+1} Ax_1^2 + b_{n+2} AX_2^2 + \dots + b_{2n} AX_n^2 = 0.
\end{aligned} \quad (16)$$

Taking dot product with  $AX_i$  ( $i = 1, \dots, n$ ) on both sides of (16) and noting that the quadrature sums in (14) are vanishing for all  $f = X^\alpha$  if  $\alpha$  has an odd component and  $|\alpha| \leq 3$ , we obtain

$$b_i AX_i \cdot AX_i = b_i \sum_{i=1}^{2n-1} a_i x_i^2 = 0, \quad i = 1, \dots, n.$$

Since the sums in the above equation are the quadrature sums of BTQF (14) for  $f(X) = X^\alpha$  with  $\alpha = 2\mathbf{e}_i$  ( $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$  being the standard basis of  $\mathbb{R}^n$ ), which are not be zero, we obtain  $b_i = 0$  for all  $i = 1, \dots, n$ . Consequently, equation (16) is reduced to

$$b_{n+1} Ax_1^2 + b_{n+2} AX_2^2 + \dots + b_{2n} AX_n^2 = 0. \quad (17)$$

Taking the dot product with  $A = (\sqrt{a_1}, \dots, \sqrt{a_{2n-1}})$  on both sides of equation (17) and noting that the quadrature sums in (14) are vanishing for all  $f = X^\alpha$  if  $\alpha = 3$ , we obtain

$$\left\| \sum_{i=1}^n \sqrt{b_{n+i}} AX_i \right\|_{\ell_2}^2 = 0.$$

Hence,

$$\sqrt{b_{n+1}} AX_1 + \dots + \sqrt{b_{2n}} AX_n = 0.$$

Similarly, we have  $b_{n+i} = 0$  for all  $n = 1, \dots, n$ . Thus, vectors (15) are linearly independent, but this is impossible because all of them have  $2n - 1$  components. This contradiction means that the number of evaluation points on the outside layer for any BTQFs over  $Sh_n$  with algebraic precision degree 3 must be more than  $2n - 1$ .

We now prove that the number of evaluation points on the inside layer for any BTQFs over  $Sh_n$  with precision degree 3 cannot be less than 2. Otherwise, if there is none or there is only one evaluation point,  $X_0 = (x_{0,1}, x_{0,2}, \dots, x_{0,n})$ , on the inside layer of  $Sh_n$ , then a BTQF over  $Sh_n$  with algebraic precision degree 3 is not exact for quadratic polynomial  $f(X) = \sum_{i=1}^n x_i^2 - b^2$  or for a cubic polynomial

$$f(X) = \left( \sum_{i=1}^n x_i^2 - b^2 \right) (x_j - x_{0,j}),$$

where  $x_{0,j} \neq 0$ . This completes the proof of theorem.  $\square$

A similar argument of the proof of Theorem 5 can be applied to solve other minimum evaluation point problem. For instance, we have the following result.

**Theorem 6** *The minimum number of the evaluation points needed for constructing a quadrature formula over an axially symmetric region in  $\mathbb{R}^n$  with algebraic precision degree 3 is  $2n$ .*

The construction of a quadrature formula of this type can be found in Section 3.9 of Stroud [33].

The minimum number of the evaluation points needed for constructing a quadrature formula over an axially symmetric region in  $\mathbb{R}^n$  with certain algebraic precision degree is topologically invariant under a reflection group action.

### 3 BTQFs with derivatives

To improve the algebraic precision degrees of BTQR's, we use the derivatives of the integrands. As examples, we will construct symmetric quadrature formulas over the surfaces of the regions  $C_2 = [-1, 1]^2$ ,  $C_3 = [-1, 1]^3$ , and the  $n$ -dimensional sphere  $S_n$ .

**Example 4.** Denote the sets of fully symmetric points  $XI = \{(1, 1)_{FS}\}$  and  $XII = \{(1, 0)_{FS}\}$ . We construct a symmetric BTQF with precision degree 5 over  $C_2 = [-1, 1]$  as follows.

$$\begin{aligned} \int_{C_2} f(x, y) dx dy &\approx a_1 \sum f_4(XI) + a_2 \sum f_4(XII) \\ &+ a_3 [f'_x(1, 1) - f'_x(-1, -1) + f'_x(1, -1) - f'_x(-1, 1) \\ &+ f'_y(1, 1) - f'_y(-1, -1) + f'_y(-1, 1) - f'_y(1, -1)] \\ &+ a_4 [f'_x(1, 0) - f'_x(-1, 0) + f'_y(0, 1) - f'_y(0, -1)]. \end{aligned}$$

Obviously, the above quadrature formula is of precision degree 5 if it is exact for  $f(x, y) = 1$ ,  $x^2$ ,  $x^4$ , and  $x^2 y^2$ . Therefore, we obtain

$$a_1 = -\frac{1}{15}, \quad a_2 = \frac{16}{15}, \quad a_3 = \frac{2}{45}, \quad a_4 = -\frac{2}{9}.$$

We use the following numerical example to show the good accuracy of the above BTQF. Considering function  $f(x, y) = e^{-x^2 - y^2}$  and applying the last quadrature to the integral of  $f(x, y)$  over  $[0, 2]^2$ , we obtain

$$\begin{aligned} \int_{[0, 2]^2} f(x, y) dx dy &= \frac{1}{4} \int_{C_2} e^{-((x+1)^2 + (y+1)^2)/4} dx dy \\ &\approx -\frac{1}{60} (e^{-2} + 2e^{-1} + 1) + \frac{4}{15} (2e^{-5/4} + 2e^{-1/4}) \\ &\quad - \frac{1}{90} (2e^{-2} + 2e^{-1}) + \frac{1}{9} e^{-5/4} = 0.5576, \end{aligned}$$

while the actual integral value is 0.5577.

Similarly, we can construct a BTQF over  $C_3 = [-1, 1]^3$  with precision degree 7 and 50 fully symmetric evaluation points  $XIII = \{(1, 1, 1)_{FS}\}$ ,  $XIV = \{(1, 0, 0)_{FS}\}$ ,  $XV = \{(1, \frac{1}{2}, 0)_{FS}\}$ , and  $XVI = \{(1, 1, 0)_{FS}\}$  as follows.

$$\begin{aligned} \int_{C_3} f(x, y, z) dx dy dz &\approx a_1 \sum f_8(XIII) + a_2 \sum f_6(XIV) \\ &+ a_3 \sum f_{24}(XV) + a_4 \sum f_{12}(XVI) + a_5 M_1 + a_6 M_2 + a_7 M_3, \end{aligned}$$

where  $a_1 = \frac{1}{5}$ ,  $a_2 = -\frac{16}{105}$ ,  $a_3 = \frac{512}{945}$ ,  $a_4 = -\frac{64}{135}$ ,  $a_5 = -\frac{11}{405}$ ,  $a_6 = -\frac{16}{81}$ ,  $a_7 = \frac{172}{2835}$ ,

$$\begin{aligned} M_1 &= f'_x(1, 1, 1) - f'_x(-1, -1, -1) + f'_x(1, 1, -1) - f'_x(-1, -1, 1) \\ &+ f'_x(1, -1, -1) - f'_x(-1, 1, 1) + f'_x(1, -1, 1) - f'_x(-1, 1, -1) \\ &+ f'_y(-1, 1, -1) - f'_y(1, -1, 1) + f'_y(-1, 1, 1) - f'_y(1, -1, -1) \\ &+ f'_y(1, 1, -1) - f'_y(-1, -1, 1) + f'_y(1, 1, 1) - f'_y(-1, -1, -1) \\ &+ f'_z(-1, 1, 1) - f'_z(1, -1, -1) + f'_z(1, -1, 1) - f'_z(-1, 1, -1) \\ &+ f'_z(1, 1, 1) - f'_z(-1, -1, -1) + f'_z(-1, -1, 1) - f'_z(1, 1, -1), \end{aligned}$$

$$M_2 = f'_x(1, 0, 0) - f'_x(-1, 0, 0) + f'_y(0, 1, 0) - f'_y(0, -1, 0) + f'_z(0, 0, 1) - f'_z(0, 0, -1),$$

and

$$\begin{aligned} M_3 = & f'_x(1, 1, 0) - f'_x(-1, -1, 0) + f'_x(1, -1, 0) - f'_x(-1, 1, 0) \\ & + f'_x(1, 0, 1) - f'_x(-1, 0, -1) + f'_x(1, 0, -1) - f'_x(-1, 0, 1) \\ & + f'_y(0, 1, 1) - f'_y(0, -1, -1) + f'_y(0, 1, -1) - f'_y(0, -1, 1) \\ & + f'_y(1, 1, 0) - f'_y(-1, -1, 0) + f'_y(-1, 1, 0) - f'_y(1, -1, 0) \\ & + f'_z(1, 0, 1) - f'_z(-1, 0, -1) + f'_z(-1, 0, 1) - f'_z(1, 0, -1) \\ & + f'_z(0, 1, 1) - f'_z(0, -1, -1) + f'_z(0, -1, 1) - f'_z(0, 1, -1). \end{aligned}$$

**Example 5.** Choose  $2n$  fully symmetric evaluation points  $XVII = \{(r, 0, \dots, 0)_{FS}\}$ . We can obtain a BTQF over  $S_n(\sum_{i=1}^n x_i^2 \leq r^2)$  with the precision degree 3 as follows.

$$\begin{aligned} \int_{S_n} f(X) dX &\approx \frac{\pi^{n/2} r^{n+1}}{2n(n+2)\Gamma(\frac{n}{2}+1)} \left[ \frac{n+2}{r} \sum f_{2n}(XVII) \right. \\ &\quad \left. - f'_{x_1}(r, 0, \dots, 0) + f'_{x_1}(-r, 0, \dots, 0) - \dots - f'_{x_n}(0, \dots, 0, r) + f'_{x_n}(0, \dots, 0, -r) \right]. \end{aligned}$$

At the end of this section, we discuss the construction of the numerical quadrature formulas over  $\bar{S}_n = \{X \in \mathbb{R}^n | X| = 1\}$  using some recent results in [35], where  $\bar{S}_n$  is the surface of the unit sphere  $B_n = B_n(1) = \{X \in \mathbb{R}^n | X| \leq 1\}$  in  $\mathbb{R}^n$ . Let  $H$  be a function defined on  $\mathbb{R}^n$  that is symmetric with respect to  $x_n$ ; i.e.,  $H(X, x_n) = H(X, -x_n)$ ,  $X \in \mathbb{R}^{n-1}$ . Then for any continuous function  $f$  defined on  $\bar{S}_n$ ,

$$\begin{aligned} & \int_{\bar{S}_n} f(Y) H(Y) d\mu_n \\ &= \int_{B_{n-1}} \left[ f\left(X, \sqrt{1-|X|^2}\right) + f\left(X, -\sqrt{1-|X|^2}\right) \right] \\ & \quad \times H(X, \sqrt{1-|X|^2}) \frac{dX}{\sqrt{1-|X|^2}}, \end{aligned} \tag{18}$$

where  $Y \in \bar{S}_n$ ,  $X \in \mathbb{R}^{n-1}$ ,  $-1 \leq t \leq 1$ , and  $d\omega_n$  is the surface measure on  $\bar{S}_n$ . The volume of  $\bar{S}_n$  is  $\omega_n = \int_{\bar{S}_n} d\mu_n = 2\pi^{n/2}/\Gamma(\frac{n}{2})$ . Formula (18), shown in Xu [35], can be proved straightforwardly by substituting  $d\mu_n = (1-t^2)^{(n-3)/2} dt d\mu_{n-1}$  and  $Y = (\sqrt{1-t^2}X, t)$  into the left-hand integral of the equation.

(18) changes a boundary integral into an integral over the interior of the boundary. Hence it can be used to derive a BTQF over  $B_n$  from a quadrature formula of an integral over  $B_{n-1}$ . Following [35], suppose that there is a quadrature formula of precision degree  $m$  on  $B_{n-1}$

$$\int_{B_{n-1}} g(X) H\left(X, \sqrt{1-|X|^2}\right) \frac{dX}{\sqrt{1-|X|^2}} \approx \sum_{i=1}^N a_i g(X_i);$$

that is, the quadrature formula is exact for all polynomials in  $\pi_m^{n-1}$ , which denotes the set of all polynomials defined in  $\mathbb{R}^{n-1}$  with a total degree not more than  $m$ .



Then there is a quadrature formula of homogeneous precision degree  $m$  on  $\bar{S}_n$ :

$$\int_{\bar{S}_n} f(Y)H(Y)d\mu_n \approx \sum_{i=1}^N a_i \left[ f\left(X_i, \sqrt{1-|X_i|^2}\right) + f\left(X_i, -\sqrt{1-|X_i|^2}\right) \right]. \quad (19)$$

Recently, Mhaskar, Narcowich, and Ward (see[28]) developed a new method for obtaining quadrature formulas on  $\bar{S}_n$ , which can be applied to the right-hand integrals of equation (20) in the following theorem, so that the BTQFs over  $B_n$  can be constructed.

**Theorem 7** *Suppose that  $F(X)$  is a continuous function defined on the sphere  $B_n(x_1^2 + \dots + x_n^2 \leq 1)$  that has  $2m$  order continuous partial derivative with respect to  $x_n$ . Then there exists the following expansion that has  $m$  terms and possesses degree  $2m - 1$  of algebraic precision.*

$$\int_{B_n} F(X)dV = \sum_{k=0}^{m-1} \frac{(-1)^k}{m!} \int_{S_{n-1}} L_k(F(X), U_m(X)) dS + \rho_m, \quad (20)$$

where  $L_k(\cdot, \cdot)$  is defined by

$$L_k(F, G) \equiv \left( \frac{\partial^k F}{\partial x_n^k} \right) \left( \frac{\partial^{m-k-1} G}{\partial x_n^{m-k-1}} \right) \left( \frac{\partial x_n}{\partial \nu} \right)$$

and  $\rho_m$  has estimate

$$|\rho_m| \leq \frac{\pi^{\frac{n}{2}} \cdot m!}{\Gamma\left(m + \frac{n}{2} + 1\right) (2m)!} \left\| \frac{\partial^{2m} F}{\partial x_n^{2m}} \right\|_C \quad (21)$$

or

$$|\rho_m| \leq \left( \frac{\pi^{\frac{n}{2}} \cdot m!}{\Gamma\left(m + \frac{n}{2} + 1\right) (2m)!} \right)^{\frac{1}{2}} \left\| \frac{\partial^{2m} F}{\partial x_n^{2m}} \right\|_{L_2}. \quad (22)$$

Formula (20) can be proved using the Green's formula successively, and it is omitted here.

## References

- [1] B.L. Burrows, *A new approach to numerical integration*, J. Inst. Math. Applies, 26(1980), 151-173.
- [2] B. Cipra, *What's Happening in the Mathematical Sciences*, American Mathematical Society, Providence, RI, 1996.
- [3] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

- [4] P.J. Davis, *Double integrals expressed as single integrals or interpolatory functions*, J. Appro. Theory, 5(1972), 276-307.
- [5] P.J. Davis and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, New York, 1975.
- [6] J.D. Federenko, *A formula for the approximate evaluation of double integrals*, Dopovidi. Akad. Nauk Ukrain. RSR, (1964), 1000-1005.
- [7] A. Ghizzetti and A. Ossicini, *Quadrature Formula*, Academic Press, New York, 1970.
- [8] W. Gröbner, *Über die Konstruktion von Systemen orthogonaler Polynome in ein-und-zwei dimensionalen Bereiche*, Monatsh. Math., 52(1948), 48-54.
- [9] T.X. He, *Boundary-type quadrature formulas without derivative terms*, J. Math. Res. Expo., 2(1981), 93-102.
- [10] T.X. He, *On the algebraic method for constructing the boundary-type quadrature formulas*, Comp. Math. (China), (1985), No.1, 1-5.
- [11] T.X. He, *Spline interpolation and its wavelet analysis*, Proceedings of the Eighth International Conference on Approximation Theory, C.K. Chui and L.L. Schumaker (eds.), World Scientific Publishing Co., Inc., 1995, 143-150.
- [12] T.X. He, *Construction of boundary quadrature formulas using wavelets*, Wavelet Applications in Signal and Image Processing III, SPIE-The International Society for Optical Engineering, A.F. Laine and M.A. Unser (eds.) 1995, 825-836.
- [13] T.X. He, *Short time Fourier transform, integral wavelet transform, and wavelet functions associated with splines*, J. Math. Anal. & Appl., 224(1998), 182-200.
- [14] T.X. He, *Boundary quadrature formulas and their applications*, Handbook of Analytic-Computational Methods in Applied Mathematics, G. Anastassiou (ed.), Chapman & Hall/CRC, New York, 2000, 773-800.
- [15] T.X. He, *Dimensionality Reducing Expansion of Multivariate Integration*, Birkhäuser, Boston, March, 2001.
- [16] E. Hernández and G. Weiss, *A First Course on Wavelets*, CRC Press, New York, 1996.
- [17] L.C. Hsu, *On a method for expanding multiple integrals in terms of integrals in lower dimensions*, Acta. Math. Acad. Sci. Hung., 14(1963). 359-367.
- [18] L.C. Hsu and T.X. He, *On the minimum estimation of the remainders in dimensionality lowering expansions with algebraic precision*, J. Math. (Wuhan), 2,3(1982), 247-255.
- [19] L.C. Hsu and Y.S. Zhou, *Numerical integration in high dimensions*, Computational Methods Series. Science Press, Beijing, 1980.

- [20] L.C. Hsu and Y.S. Zhou, *Two classes of boundary type cubature formulas with algebraic precision*, *Calcolo*, 23(1986), 227-248.
- [21] D.V. Ionescu, *Generalization of a quadrature formula of N. Obreschkoff for double integrals* (Romanian), *Stud. Cerc. Mat.*, 17(1965), 831-841.
- [22] E. G. Kalnins, W. Miller Jr., and M. V. Tratnik, Families of orthogonal and biorthogonal polynomials on the  $N$ -sphere. *SIAM J. Math. Anal.* 22 (1991), no. 1, 272-294.
- [23] P. Keast and J.C. Diaz, *Fully symmetric integration formula for the surface of the sphere in  $S$  dimension*, *SIAM J. Numer. Anal.* 20(1983), 406-419.
- [24] L.J. Kratz, *Replacing a double integral with a single integral*, *J. Appro. Theory*, 27(1979), 379-390.
- [25] V. I. Lebedev and A. L. Skorokhodov, Quadrature formulas for a sphere of orders 41, 47 and 53. (Russian) *Dokl. Akad. Nauk* 324 (1992), no. 3, 519-524; translation in *Russian Acad. Sci. Dokl. Math.* 45 (1992), no. 3, 587-592
- [26] M. Levin, *On a method of evaluating double integrals*, *Tartu Riikl. Ül. Toime-tised*, 102(1961), 338-341.
- [27] M. Levin, *Extremal problems connected with a quadrature formula*, *Eesti NSV Tead. Akad. Toimetised Füüs-Mat. Tehn. Seer.*, 12(1963), 44-56.
- [28] H.N. Mhaskar, F.J. Narcowich, and J.D. Ward, *Quadrature formulas on spheres using scattered data*, preprint, 2000.
- [29] N. Obreschkoff, *Neue Quadraturformeln*, *Abhandl. d. preuss. Akad. d. Wiss., Math. Natur. wiss. Kl.*, 4(1940), 1-20.
- [30] M. Sadowsky, *A formula for the approximate computation of a triple integral*, *Amer. Math. Monthly*, 47(1940), 539-543.
- [31] D.D. Stancu, *Sur quelques formules generales de quadrature du type Gauss-Christoffel*, *Mathematica (Cluj)*, 1(1959), 167-182.
- [32] D. D. Stancu and A.H. Stroud, Quadrature formulas with simple Gaussian nodes and multiple fixed nodes, *Math. Comp.*, 17(1963), 384-394.
- [33] A.H. Stroud, *Approximate Calculation of Multiple Integrals*, Prentice-Hall, Englewood Cliffs, N.H., 1971.
- [34] G.G. Walter, *Wavelets and Other Orthogonal Systems with Applications*, CRC Press, Ann Arbor, 1994.
- [35] Y. Xu, *Orthogonal polynomials and cubature formulae on spheres and on balls*, *SIAM J. Math. Anal.*, 29(1998), 779-793.
- [36] Y.S. Zhou and T.X. He, Higher dimensional Korkin Theorem,

# Theoretical Analysis and Numerical Realization of Bioluminescence Tomography

Rongfang Gong<sup>1</sup>, Xiaoliang Cheng<sup>2</sup> and Weimin Han<sup>3</sup>

**Abstract.** Mathematically, bioluminescence tomography (BLT) is an ill-posed inverse source problem. In this paper, a new formulation for the BLT problem is developed. It is used to explain rigorously the reason behind the loss of the continuous dependence of the light source function solution on the measurements. On the basis of the formulation, an approximation using finite element method is provided. Some error estimates for the reconstructed light source function are obtained. By using adjoint equations, a simple but efficient iterative scheme is explored. Several numerical examples are presented to test the performance of the formulation and the iterative scheme.

**Keywords.** Bioluminescence tomography, ill-posed problem, finite element method

## 1 Introduction

The function of molecular imaging is to help study biological processes *in vivo* at the cellular and molecular levels, see e.g. [8, 25, 29, 31, 32]. It may non-invasively differentiate normal from diseased conditions. While some classic techniques do reveal information on micro-structures of the tissues, only recently have molecular probes been developed along with associated imaging technologies that are sensitive and specific for detecting molecular targets in animals and humans. A molecular probe has a high affinity for attaching itself to a target molecule and a tagging ability with a marker molecule that can be tracked outside a living body. Molecular imaging contains a lot of modalities, see [7] for detail. Among them, optical imaging, especially fluorescence and bioluminescence imaging, has attracted remarkable attention for its unique advantages regarding performance and cost-effectiveness.

Bioluminescence tomography (BLT) [12, 33, 34] is an emerging and promising bioluminescence imaging modality. The major issue of BLT is the determination of the distribution of *in vivo* bioluminescent source. In BLT, we reconstruct an internal bioluminescent source from the measured bioluminescent signal on the external surface of a small animal. The problem of determining the photon density on the small animal surface from the bioluminescent

---

<sup>1</sup>Department of Mathematics, Zhejiang University, Hangzhou 310027, P.R. China. E-mail: gongrongfang319@yahoo.com.cn

<sup>2</sup>Department of Mathematics, Zhejiang University, Hangzhou 310027, P.R. China. E-mail: xiaoliangcheng@zju.edu.cn

<sup>3</sup>Department of Mathematics, University of Iowa, Iowa City, IA 52242, U.S.A. E-mail: whan@math.uiowa.edu

source distribution within the animal requires accurate representation of photon transport in biological tissue. Photon propagation in biological tissue is governed by the radiative transfer equation (RTE) [24]. However, the RTE is highly dimensional and presents a serious challenge for its accurate numerical simulations given the current level of development in computer software and hardware. Because the mean-free path of the photon is between 500 nm and 1000 nm in biological tissues, which is very small compared to the size of a typical object in this context, the predominant phenomenon is scattering. Usually a diffusion approximation of the RTE is employed [1, 30].

Based on the diffusion approximation equation, many theoretical analysis and numerical methods are explored, see e.g. [6, 7, 10, 18, 27]. To improve the accuracy of the reconstructed light source function, multispectral systems are developed [4, 11, 19, 21, 28]. We refer to [20, 35] for the BLT problem related to the optical properties issue. A recent survey of biomedical background, mathematical theory and numerical approximation for BLT is [22].

In this paper, we study the BLT problem through a new perspective that leads to better convergence behavior for its numerical solution.

In Section 2, a simplified version of the BLT problem is reduced to an operator equation. The operator is compact, which explains why the source function in the BLT problem does not depend continuously on measurements.

In Section 3, based on the discussion in Section 2, a new Tikhonov-type regularized formulation is provided to compute the light source function. In the formulation, the measurements on the boundary are transferred to a knowledge in the problem domain, which makes the BLT problem more regular. The well-posedness of the new formulation, including the solution existence, uniqueness and continuous dependence, is shown. The limiting behavior is discussed for the regularized solution when the regularization approaches zero.

In Section 4, a finite element approximation for the BLT problem based on the new formulation is studied. Specifically, we use piecewise constants to approximate the source function and standard linear elements to approximate the state function. Error estimates are derived with improved convergence orders compared to those found in [16, 18].

In Subsection 5.1, an iterative scheme for the BLT reconstruction based on the given formulation is shown. By introducing adjoint equations, we avoid computing the inversion of an elliptic partial differential operator. Some detail on the implementation is also given. Two numerical examples are presented in Subsection 5.2 to show the numerical performance of the method discussed in this paper.

## 2 Ill-posedness of the BLT problem

We first introduce the classical formulation of the diffusion based BLT problem. Let the biological medium occupies a non-empty, open and bounded set  $\Omega \subset \mathbb{R}^d$ ,  $d \leq 3$ . The boundary  $\Gamma$  of the domain  $\Omega$  is assumed Lipschitz continuous. Denote by  $D = (3(\mu_a + \mu'_s))^{-1}$  with absorption coefficient  $\mu_a$  and reduced scattering coefficient  $\mu'_s$ , and denote by  $\partial_\nu$  the outward normal differentiation operator. Then the classical formulation of the BLT problem based on the diffusion approximation is the following ([7, 18]).

**Problem 2.1** *Given  $D > 0$ ,  $\mu_a \geq 0$ ,  $g_1$  and  $g_2$ , suitably smooth, find a bioluminescent source  $p$  such that the solution  $u$  of the boundary-value problem*

$$-\operatorname{div}(D\nabla u) + \mu_a u = p\chi_{\Omega_0} \quad \text{in } \Omega, \quad (2.1)$$

$$D\partial_\nu u = g_2 \quad \text{on } \Gamma \quad (2.2)$$

*satisfies*

$$u = g_1 \quad \text{on } \Gamma. \quad (2.3)$$

Here  $\Omega_0$ , known as the permissible region, is a measurable subset of  $\Omega$ , and  $\chi_{\Omega_0}$  is the characteristic function of  $\Omega_0$ :  $\chi_{\Omega_0}(x)$  equals 1 for  $x \in \Omega_0$ , and equals 0 for  $x \in \Omega/\Omega_0$ .

It is shown in [18] that the pointwise BLT Problem 2.1 may have infinitely many solutions or may have no solution, depending on the choice of the function set where we look for the source function. It is also mentioned there even when the solution existence and uniqueness issues could be settled, the source function solution does not depend continuously on the measurement. This section is devoted to a rigorous proof of this instability statement for the BLT problem in the simplified case where the permissible region is the entire domain and the admissible set for the source function is the entire  $L^2(\Omega)$  space.

We use standard notations for Sobolev spaces ([2]). Denote  $V = H^1(\Omega)$ ,  $V_0 = H_0^1(\Omega)$ ,  $Q = L^2(\Omega_0)$ . Our basic assumptions on the data throughout the paper are:  $D \in L^\infty(\Omega)$  and  $D \geq D_0$  for some constant  $D_0 > 0$ ,  $\mu_a \in L^2(\Omega)$  and  $\mu_a \geq \mu_0$  for some constant  $\mu_0 > 0$ ,  $g_1 \in H^{1/2}(\Gamma)$ , and  $g_2 \in L^2(\Gamma)$ . We further let  $V_{g_1} = \{v \in V \mid v = g_1 \text{ on } \Gamma\}$ .

Define

$$a(u, v) = \int_{\Omega} (D\nabla u \cdot \nabla v + \mu_a u v) \, dx \quad \forall u, v \in V. \quad (2.4)$$

Then  $a(\cdot, \cdot)$  is symmetric, continuous and coercive on  $V$ . Therefore, by the Lax-Milgram Lemma ([2, 13]), for any  $q \in Q$ , the problems

$$u_D(q, g_1) \in V_{g_1}, \quad a(u_D(q, g_1), v) = (q, v)_Q \quad \forall v \in V_0 \quad (2.5)$$

and

$$u_N(q, g_2) \in V, \quad a(u_N(q, g_2), v) = (q, v)_Q + (g_2, v)_{L^2(\Gamma)} \quad \forall v \in V \quad (2.6)$$

each has a unique solution. Moreover,

$$\|u_D(q, g_1)\|_V \leq c(\|q\|_Q + \|g_1\|_{H^{1/2}(\Gamma)}), \quad (2.7)$$

$$\|u_N(q, g_2)\|_V \leq c(\|q\|_Q + \|g_2\|_{L^2(\Gamma)}). \quad (2.8)$$

We write  $u_D(q) = u_D(q, 0)$ ,  $\tilde{u}_D(g_1) = u_D(0, g_1)$ ,  $u_N(q) = u_N(q, 0)$ , and  $\tilde{u}_N(g_2) = u_N(0, g_2)$ . Then  $u_D(q, g_1) = u_D(q) + \tilde{u}_D(g_1)$  and  $u_N(q, g_2) = u_N(q) + \tilde{u}_N(g_2)$ .

Next we introduce a weak formulation for Problem 2.1 based on both Dirichlet and Neumann boundary problems (2.5) and (2.6). Define

$$\begin{aligned} s(p, q) &= (u_D(p) - u_N(p), u_D(q) - u_N(q))_{L^2(\Omega)} \quad \forall p, q \in Q, \\ l(q) &= (\tilde{u}_N(g_2) - \tilde{u}_D(g_1), u_D(q) - u_N(q))_{L^2(\Omega)} \quad \forall q \in Q. \end{aligned} \quad (2.9)$$

Apparently, the bilinear form  $s$  is symmetric and positive semi-definite over  $Q$ . Moreover, by Schwarz inequality, trace theorem [3, Theorem 1.6.6], together with the bounds (2.7) and (2.8), we conclude that both  $s$  and  $l$  are continuous on  $Q$ .

Denote by  $\mathcal{S} : Q \rightarrow Q$  the operator through the relation  $(\mathcal{S}p, q)_Q = s(p, q)$  for any  $p, q \in Q$ . Then  $\mathcal{S}$  is symmetric, positive semi-definite and continuous. Moreover, from Riesz representation theorem, there is an element denoted again by  $l \in Q$  such that  $(l, q)_Q = l(q)$  for any  $q \in Q$ .

We let  $\Omega_0 = \Omega$  in the rest of this section, and introduce a new problem.

**Problem 2.2** Find  $p \in Q$  such that

$$\mathcal{S}p = l \quad \text{in } Q. \quad (2.10)$$

We have the following result.

**Proposition 2.3** If  $p \in Q$  solves Problem 2.1, then it is a solution of Problem 2.2.

**Proof.** Let  $p_*$  be a solution of Problem 2.1. Then,  $u_D(p_*, g_1) = u_N(p_*, g_2)$ . So

$$u_D(p_*) - u_N(p_*) = \tilde{u}_N(g_2) - \tilde{u}_D(g_1).$$

Thus,

$$(u_D(p_*) - u_N(p_*), u_D(q) - u_N(q))_{L^2(\Omega)} = (\tilde{u}_N(g_2) - \tilde{u}_D(g_1), u_D(q) - u_N(q))_{L^2(\Omega)} \quad \forall q \in Q,$$

i.e.,

$$s(p_*, q) = l(q) \quad \forall q \in Q.$$

Hence,  $p_*$  is a solution of Problem 2.2. ■

It is possible to demonstrate the equivalence between Problem 2.1 and Problem 2.2. For this purpose, we introduce the following stronger smoothness assumptions on the data:

$$\Gamma \in C^{1,1}, \quad D \in C^{0,1}(\overline{\Omega}), \quad g_1 \in H^{3/2}(\Gamma), \quad g_2 \in H^{1/2}(\Gamma). \quad (2.11)$$

We recall the following result ([18]).

**Lemma 2.4** *Let  $\Omega \subset \mathbb{R}^d$  be an open bounded subset with a  $C^{1,1}$  boundary and  $u \in H^2(\Omega)$ . Then there are infinitely many functions  $v \in H^2(\Omega)$  such that*

$$\gamma v = \gamma u, \quad \gamma \partial_\nu v = \gamma \partial_\nu u.$$

Here  $\gamma$  stands for the trace operator.

We have the following converse of Proposition 2.3.

**Proposition 2.5** *Assume (2.11). Then a solution of Problem 2.2 solves Problem 2.1.*

**Proof.** Let  $p_* \in Q$  be a solution of Problem 2.2:

$$(\mathcal{S}p_*, q)_Q = (l, q)_Q \quad \forall q \in Q$$

or equivalently

$$(u_D(p_*, g_1) - u_N(p_*, g_2), u_D(q) - u_N(q))_{L^2(\Omega)} = 0 \quad \forall q \in Q. \quad (2.12)$$

Define an operator  $\mathcal{E}$  from  $H^2(\Omega)$  to  $L^2(\Omega)$ :  $\mathcal{E}u = -\operatorname{div}(D\nabla u) + \mu_a u$  for  $u \in H^2(\Omega)$ , and set  $u_* = u_D(p_*, g_1) - u_N(p_*, g_2)$ . Then  $u_* \in H^2(\Omega)$  from (4.10) and (4.11) below, and  $\mathcal{E}u_* = 0$ . Denote  $g_{1,*} = \gamma u_*$  and  $g_{2,*} = \gamma D\partial_\nu u_*$ . Then,  $g_{1,*} \in H^{3/2}(\Gamma)$  and  $g_{2,*} \in H^{1/2}(\Gamma)$ . By Lemma 2.4, there exists a function  $u_{D,*} \neq u_*$  in  $H^2(\Omega)$  such that

$$u_{D,*} = 0, \quad D\partial_\nu u_{D,*} = g_{2,*} \quad \text{on } \Gamma.$$

Let  $q_* = \mathcal{E}u_{D,*} \in Q$  and  $u_{N,*} = u_{D,*} - u_* \in H^2(\Omega)$ . Then  $u_{D,*} = u_D(q_*)$  and  $u_* = u_{D,*} - u_{N,*}$ . From

$$\mathcal{E}u_{N,*} = \mathcal{E}u_{D,*} - \mathcal{E}u_* = q_* - 0 = q_* \quad \text{in } \Omega$$

and

$$D\partial_\nu u_{N,*} = D\partial_\nu u_{D,*} - D\partial_\nu u_* = g_{2,*} - D\partial_\nu u_* = 0 \quad \text{on } \Gamma,$$



we know that  $u_{N,*} = u_N(q_*)$ , i.e., there exists a  $q_* \in Q$  such that  $u_D(q_*) - u_N(q_*) = u_*$ . Substitute  $q_*$  for  $q$  in (2.12) to give  $u_D(p_*, g_1) = u_N(p_*, g_2)$ , which shows that  $p_*$  is a solution of Problem 2.1.  $\blacksquare$

We now present a result on the compactness of the operator  $\mathcal{S}$ .

**Theorem 2.6** *The operator  $\mathcal{S} : Q \rightarrow Q$  is compact.*

**Proof.** Let  $\{p^n\}_n \subset Q$  be bounded. Then there is a subsequence, denoted again by  $\{p^n\}_n$ , which converges weakly in  $Q$  to some element  $p^* \in Q$  because of the reflexivity of space  $Q$ .

Let  $u_D^n = u_D(p^n)$ ,  $u_N^n = u_N(p^n)$ , i.e.,  $u_D^n \in V_0$ ,  $u_N^n \in V$ , and

$$a(u_D^n, v) = (p^n, v)_Q \quad \forall v \in V_0, \quad (2.13)$$

$$a(u_N^n, v) = (p^n, v)_Q \quad \forall v \in V. \quad (2.14)$$

Then  $\{u_D^n\}_n$  and  $\{u_N^n\}_n$  are bounded in  $V$  from the properties (2.7) and (2.8). Hence, we can extract two further subsequences, denoted again by  $\{u_D^n\}_n$  and  $\{u_N^n\}_n$ , which converge weakly in  $V$  and strongly in  $Q$  to  $u_D^* \in V_0$  and  $u_N^* \in V$ , respectively. Let  $n \rightarrow \infty$  in (2.13) and (2.14) to get  $u_D^* = u_D(p^*)$  and  $u_N^* = u_N(p^*)$ . Strong convergence of  $\{u_D^n\}_n$  to  $u_D^*$  in  $V$  follows from

$$\alpha \|u_D^n - u_D^*\|_V^2 \leq a(u_D^n - u_D^*, u_D^n - u_D^*) = \int_{\Omega} (p^n - p^*) (u_D^n - u_D^*) dx \rightarrow 0$$

as  $n \rightarrow \infty$ . Similarly,  $u_N^n \rightarrow u_N^*$  as  $n \rightarrow \infty$ .

Denote  $s^n = \mathcal{S}p^n$ . Then  $\{s^n\}_n$  is bounded in  $Q$ . Repeating the above argument, we conclude that there exists an element  $s^* \in Q$  such that

$$s^n \rightharpoonup s^* \text{ in } Q, \quad u_D(s^n) \rightarrow u_D(s^*), \quad u_N(s^n) \rightarrow u_N(s^*) \text{ in } V \text{ as } n \rightarrow \infty.$$

Since

$$(s^n, q)_Q = (\mathcal{S}p^n, q)_Q = s(p^n, q) = (u_D^n - u_N^n, u_D(q) - u_N(q))_{L^2(\Omega)} \quad \forall q \in Q,$$

we have  $s^* = \mathcal{S}p^*$  by letting  $n \rightarrow \infty$ . Consequently, strong convergence of  $s^n$  to  $s^*$  in  $Q$  follows from

$$\begin{aligned} \|s^n - s^*\|_Q^2 &= (\mathcal{S}p^n - \mathcal{S}p^*, s^n - s^*)_Q \\ &= s(p^n - p^*, s^n - s^*) \\ &= (u_D^n - u_D^* - (u_N^n - u_N^*), u_D(s^n) - u_D(s^*) - (u_N(s^n) - u_N(s^*)))_{L^2(\Omega)} \\ &\leq (\|u_D^n - u_D^*\|_V + \|u_N^n - u_N^*\|_V) (\|u_D(s^n) - u_D(s^*)\|_V + \|u_N(s^n) - u_N(s^*)\|_V) \\ &\rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ , and the proof is complete.  $\blacksquare$

Compactness of  $\mathcal{S}$  explains the instability of solutions of the BLT problem with respect to the measurement data.

### 3 A new regularized reformulation for the BLT problem

In this section, a Tikhonov-type regularization method is used to reconstruct numerically the source function of the BLT problem. We seek the source function in an admissible set  $Q_{ad} \subset Q$ , which is assumed to be bounded, closed and convex set. Usually, we take

$$Q_{ad} = \{q \in Q \mid q \geq 0 \text{ a.e. in } \Omega_0\}.$$

For any  $\varepsilon \geq 0$ , define a bilinear form  $s_\varepsilon(\cdot, \cdot)$  over  $Q \times Q$ :

$$s_\varepsilon(p, q) = s(p, q) + \varepsilon (p, q)_Q \quad \forall p, q \in Q. \quad (3.1)$$

Observe that  $s_\varepsilon$  is symmetric, continuous and coercive on  $Q \times Q$  for each  $\varepsilon > 0$ .

Note that Problem 2.2 is equivalent to minimizing the functional  $\frac{1}{2}s(q, q) - l(q)$  over the space  $Q$ . Thus, we introduce the following regularized reformulation of the BLT problem.

**Problem 3.1** Find  $p_\varepsilon \in Q_{ad}$  such that

$$J_\varepsilon(p) = \min_{q \in Q_{ad}} J_\varepsilon(q),$$

where

$$J_\varepsilon(q) = \frac{1}{2}s_\varepsilon(q, q) - l(q).$$

Easily,

$$J_\varepsilon(q) = \frac{1}{2}\|u_D(q, g_1) - u_N(q, g_2)\|_{L^2(\Omega)}^2 - \frac{1}{2}\|\tilde{u}_D(g_1) - \tilde{u}_N(g_2)\|_{L^2(\Omega)}^2 + \frac{\varepsilon}{2}\|q\|_Q^2.$$

We note that Problem 3.1 is equivalent to finding  $p_\varepsilon \in Q_{ad}$  such that

$$s_\varepsilon(p_\varepsilon, q - p_\varepsilon) \geq l(q - p_\varepsilon) \quad \forall q \in Q_{ad}. \quad (3.2)$$

If  $Q_{ad}$  is a subspace of  $Q$ , (3.2) reduces to

$$s_\varepsilon(p_\varepsilon, q) = l(q) \quad \forall q \in Q_{ad}.$$

Similar to  $\mathcal{S}$ , we denote by  $\mathcal{S}_\varepsilon : Q \rightarrow Q$  the operator such that  $(\mathcal{S}_\varepsilon p, q)_Q = s_\varepsilon(p, q)$  for any  $p, q \in Q$ . Then  $\mathcal{S}_\varepsilon = \mathcal{S} + \varepsilon \mathcal{I}$  is symmetric, continuous and coercive for each  $\varepsilon > 0$ , where  $\mathcal{I}$  stands for the identity operator over  $Q$ .

Next we discuss the well-posedness of Problem 3.1. Standard arguments ([2, 26]) lead to the next result.

**Proposition 3.2** *Problem 3.1 admits a unique solution  $p_\varepsilon$  in  $Q$ .*

Regarding solution stability of Problem 3.1, we have the following result.

**Theorem 3.3** *The solution  $p_\varepsilon$  of Problem 3.1 depends continuously on  $\varepsilon > 0$ ,  $D \in L^\infty(\Omega)$ ,  $\mu_a \in L^\infty(\Omega)$ ,  $g_1 \in H^{1/2}(\Gamma)$  and  $g_2 \in L^2(\Gamma)$ .*

To prove Theorem 3.3, we need some preparations. Let  $a(\cdot, \cdot)$  be given in (2.4). Assume  $D^\delta \in L^\infty(\Omega)$  and  $\mu_a^\delta \in L^\infty(\Omega)$  such that  $\|D^\delta\|_{L^\infty(\Omega)} \leq \delta D_0$  and  $\|\mu_a^\delta\|_{L^\infty(\Omega)} \leq \delta \mu_0$  for a small positive constant  $\delta$  to be specified below. For any  $q \in Q$ , let  $u_D^\delta(q, g_1) \in V_{g_1}$  and  $u_N^\delta(q, g_2) \in V$  be the unique solutions of the problems

$$a^\delta(u, v) = (q, v)_Q \quad \forall v \in V_0, \quad (3.3)$$

$$a^\delta(u, v) = (q, v)_Q + (g_2, v)_{L^2(\Gamma)} \quad \forall v \in V, \quad (3.4)$$

respectively, where

$$a^\delta(u, v) = \int_{\Omega} ((D + D^\delta) \nabla u \nabla v + (\mu_a + \mu_a^\delta) u v) dx \quad (3.5)$$

and  $\delta < 1$ . Again, we rewrite  $u_D^\delta(q)$ ,  $\tilde{u}_D^\delta(g_1)$ ,  $u_N^\delta(q)$  and  $\tilde{u}_N^\delta(g_2)$  for  $u_D^\delta(q, 0)$ ,  $u_D^\delta(0, g_1)$ ,  $u_N^\delta(q, 0)$  and  $u_N^\delta(0, g_2)$  respectively. Then we have the following estimates.

**Lemma 3.4** *For a properly small  $\delta > 0$  and any  $q \in Q$ , there exists a constant  $c$  such that*

$$\begin{aligned} \|u_D^\delta(q) - u_D(q)\|_V &\leq c \delta \|q\|_Q, \\ \|u_N^\delta(q) - u_N(q)\|_V &\leq c \delta \|q\|_Q, \\ \|\tilde{u}_D^\delta(g_1) - \tilde{u}_D(g_1)\|_V &\leq c \delta \|g_1\|_{H^{1/2}(\Gamma)}, \\ \|\tilde{u}_N^\delta(g_2) - \tilde{u}_N(g_2)\|_V &\leq c \delta \|g_2\|_{L^2(\Gamma)}. \end{aligned}$$

**Proof.** We prove the first estimate, the others can be verified similarly. We recall that

$$a(u_D(q), v) = (q, v)_Q \quad \forall v \in V_0. \quad (3.6)$$

Subtract (3.6) from (3.3) to get

$$a(u_D^\delta(q) - u_D(q), v) = - \int_{\Omega} (D^\delta \nabla u_D^\delta(q) \nabla v + \mu_a^\delta u_D^\delta(q) v) dx \quad \forall v \in V_0.$$

Because  $u_D^\delta(q) - u_D(q) \in V_0$  for any  $q \in Q$ , we can take  $v$  for  $u_D^\delta(q) - u_D(q)$  in above equation. Then by using of the coercivity of bilinear form  $a(\cdot, \cdot)$  and Schwarz inequality, we obtain

$$\|u_D^\delta(q) - u_D(q)\|_V \leq c \delta \|u_D^\delta(q)\|_V. \quad (3.7)$$

For  $\delta < 1$ , from the regularity property (2.7), we have

$$\|u_D^\delta(q)\|_V \leq c \|q\|_Q$$

which reduces (3.7) to

$$\|u_D^\delta(q) - u_D(q)\|_V \leq c \delta \|q\|_Q.$$

Therefore, we complete the proof. ■

Let  $\varepsilon^\delta \in \mathbb{R}$ ,  $g_1^\delta \in H^{1/2}(\Gamma)$  and  $g_2^\delta \in L^2(\Gamma)$  be such that  $|\varepsilon^\delta| \leq \varepsilon \delta$ ,  $\|g_1^\delta\|_{H^{1/2}(\Gamma)} \leq \delta \|g_1\|_{H^{1/2}(\Gamma)}$  and  $\|g_2^\delta\|_{L^2(\Gamma)} \leq \delta \|g_2\|_{L^2(\Gamma)}$ , and define bilinear form  $s_\varepsilon^\delta(\cdot, \cdot)$  and linear functional  $l^\delta(\cdot)$  by

$$\begin{aligned} s_\varepsilon^\delta(p, q) &= (u_D^\delta(p) - u_N^\delta(p), u_D^\delta(q) - u_N^\delta(q))_{L^2(\Omega)} + (\varepsilon + \varepsilon^\delta)(p, q)_Q \quad \forall p, q \in Q, \\ l^\delta(q) &= (\tilde{u}_N^\delta(g_2 + g_2^\delta) - \tilde{u}_D^\delta(g_1 + g_1^\delta), u_D^\delta(q) - u_N^\delta(q))_{L^2(\Omega)} \quad \forall q \in Q. \end{aligned}$$

Note that  $\tilde{u}_N^\delta(g_2 + g_2^\delta) = \tilde{u}_N^\delta(g_2) + \tilde{u}_N^\delta(g_2^\delta)$  and  $\tilde{u}_D^\delta(g_1 + g_1^\delta) = \tilde{u}_D^\delta(g_1) + \tilde{u}_D^\delta(g_1^\delta)$ . Denote by  $p_\varepsilon^\delta$  the solution of Problem 3.1 with  $s_\varepsilon$  and  $l$  replaced by  $s_\varepsilon^\delta$  and  $l^\delta$  respectively. For  $0 \leq \delta < 1$ ,  $s_\varepsilon^\delta$  is symmetric, bounded and coercive on  $Q$ , and  $l^\delta$  is continuous on  $Q$ . Hence, for  $0 \leq \delta < 1$ ,  $p_\varepsilon^\delta$  uniquely exists. Set  $\Delta s^\delta(p, q) = s_\varepsilon^\delta(p, q) - s_\varepsilon(p, q)$  and  $\Delta l^\delta(q) = l^\delta(q) - l(q)$ . Note that  $\Delta s^\delta(p, q)$  is independent of  $\varepsilon$ . We have the following bounds on  $\Delta s^\delta$  and  $\Delta l^\delta$ .

**Lemma 3.5** *There is a constant  $c > 0$  such that for any  $p, q \in Q$  and  $0 \leq \delta < 1$ ,*

$$|\Delta s^\delta(p, q)| \leq c \delta \|p\|_Q \|q\|_Q, \quad (3.8)$$

$$|\Delta l^\delta(q)| \leq c \delta (\|g_1\|_{H^{1/2}(\Gamma)} + \|g_2\|_{L^2(\Gamma)}) \|q\|_Q. \quad (3.9)$$

**Proof.** From definitions of  $s_\varepsilon$  and  $s_\varepsilon^\delta$ , for any  $p, q \in Q$ , we have

$$\begin{aligned} \Delta s^\delta(p, q) &= (u_D^\delta(p) - u_D(p) - (u_N^\delta(p) - u_N(p)), u_D^\delta(q) - u_N^\delta(q))_{L^2(\Omega)} \\ &\quad + (u_D(p) - u_N(p), u_D^\delta(q) - u_D(q) - (u_N^\delta(q) - u_N(q)))_{L^2(\Omega)} + \varepsilon^\delta (p, q)_Q, \\ \Delta l^\delta(q) &= (\tilde{u}_N^\delta(g_2) - \tilde{u}_N(g_2) - (\tilde{u}_D^\delta(g_1) - \tilde{u}_D(g_1)), u_D^\delta(q) - u_N^\delta(q))_{L^2(\Omega)} \\ &\quad + (\tilde{u}_N(g_2) - \tilde{u}_D(g_1), u_D^\delta(q) - u_D(q) - (u_N^\delta(q) - u_N(q)))_{L^2(\Omega)} \\ &\quad + (\tilde{u}_N^\delta(g_2^\delta) - \tilde{u}_D^\delta(g_1^\delta), u_D^\delta(q) - u_N^\delta(q))_{L^2(\Omega)}. \end{aligned}$$

Consequently, (3.8) and (3.9) follow immediately from Schwarz inequality, Lemma 3.4, and the regularity properties (2.7), (2.8). The proof is complete. ■

**Proof of Theorem 3.3.** Now we can prove the stability Theorem 3.3. We recall that  $p_\varepsilon \in Q_{ad}$  and  $p_\varepsilon^\delta \in Q_{ad}$  are such that

$$s_\varepsilon(p_\varepsilon, q - p_\varepsilon) \geq l(q - p_\varepsilon) \quad \forall q \in Q_{ad} \quad (3.10)$$

and

$$s_\varepsilon^\delta(p_\varepsilon^\delta, q - p_\varepsilon^\delta) \geq l^\delta(q - p_\varepsilon^\delta) \quad \forall q \in Q_{ad} \quad (3.11)$$

respectively. Replace  $q = p_\varepsilon^\delta$  in (3.10) and  $q = p_\varepsilon$  in (3.11) and add them to get

$$\begin{aligned} s_\varepsilon(p_\varepsilon^\delta - p_\varepsilon, p_\varepsilon^\delta - p_\varepsilon) &\leq -(s_\varepsilon^\delta(p_\varepsilon^\delta - p_\varepsilon, p_\varepsilon^\delta) - s_\varepsilon(p_\varepsilon^\delta - p_\varepsilon, p_\varepsilon^\delta)) + l^\delta(p_\varepsilon^\delta - p_\varepsilon) - l(p_\varepsilon^\delta - p_\varepsilon) \\ &\equiv -\Delta s^\delta(p_\varepsilon^\delta - p_\varepsilon, p_\varepsilon^\delta) + \Delta l^\delta(p_\varepsilon^\delta - p_\varepsilon). \end{aligned}$$

From the coercivity of  $s_\varepsilon$  for  $\varepsilon > 0$  together with Lemma 3.5, we have

$$\begin{aligned} \alpha(\varepsilon)\|p_\varepsilon^\delta - p_\varepsilon\|_Q^2 &\leq s_\varepsilon(p_\varepsilon^\delta - p_\varepsilon, p_\varepsilon^\delta - p_\varepsilon) \\ &\leq |\Delta s^\delta(p_\varepsilon^\delta - p_\varepsilon, p_\varepsilon^\delta)| + |\Delta l^\delta(p_\varepsilon^\delta - p_\varepsilon)| \\ &\leq c\delta\|p_\varepsilon^\delta - p_\varepsilon\|_Q (\|p_\varepsilon^\delta\|_Q + \|g_1\|_{H^{1/2}(\Gamma)} + \|g_2\|_{L^2(\Gamma)}) \\ &\leq c\delta\|p_\varepsilon^\delta - p_\varepsilon\|_Q (\|p_\varepsilon^\delta - p_\varepsilon\|_Q + \|p_\varepsilon\|_Q + \|g_1\|_{H^{1/2}(\Gamma)} + \|g_2\|_{L^2(\Gamma)}) \end{aligned}$$

or

$$(\alpha(\varepsilon) - c\delta)\|p_\varepsilon^\delta - p_\varepsilon\|_Q \leq c\delta (\|p_\varepsilon\|_Q + \|g_1\|_{H^{1/2}(\Gamma)} + \|g_2\|_{L^2(\Gamma)}), \quad (3.12)$$

where  $\alpha(\varepsilon)$  is a positive coercivity constant for bilinear form  $s_\varepsilon$  which may depend on  $\varepsilon$ . For a small enough  $\delta$ , there is a positive constant  $c(\varepsilon)$  independent of  $\delta$  such that  $\alpha(\varepsilon) - c\delta \geq c(\varepsilon)$ . Then (3.12) reduces to

$$\|p_\varepsilon^\delta - p_\varepsilon\|_Q \leq c\delta (\|p_\varepsilon\|_Q + \|g_1\|_{H^{1/2}(\Gamma)} + \|g_2\|_{L^2(\Gamma)})$$

which shows the convergence  $p_\varepsilon^\delta \rightarrow p_\varepsilon$  in  $Q$  when  $\delta \rightarrow 0$ .  $\square$

We now explore the limit behavior of the solution of Problem 3.1 as the regularization parameter  $\varepsilon \rightarrow 0$ . Denote by  $\mathcal{Z}$  the solution set of Problem 3.1 with  $\varepsilon = 0$ . Then if it is non-empty,  $\mathcal{Z}$  is a closed, convex subset of space  $Q$ . Denote by  $p_*$  the unique element in  $\mathcal{Z}$  with minimal  $Q$  norm, that is,

$$\|p_*\|_Q = \inf_{p \in \mathcal{Z}} \|p\|_Q. \quad (3.13)$$

We have the following convergence result; its proof is similar to that of [18, Proposition 3.5].

**Theorem 3.6** *Assume the solution set  $\mathcal{Z}$  is nonempty. Then*

$$p_\varepsilon \rightarrow p_* \text{ in } Q, \quad \text{as } \varepsilon \rightarrow 0.$$

## 4 A finite element approximation

In this section, we consider the problem of approximating the solution of Problem 3.1. Standard finite element method (FEM) is applied to discretize this problem.

We use constant finite element space for an approximation of the light source space  $Q$ . Specifically, let  $\{\mathcal{T}_{0,H}\}_H$  be a regular family of triangulations over domains  $\overline{\Omega}_0 \subset \overline{\Omega}$  with meshsize  $H > 0$ . For each triangulation  $\mathcal{T}_{0,H} = \{K_H\}$ , define finite element space  $Q^H = \{T \in Q \mid T|_{K_H} \in \mathcal{P}_0(K), \forall K_H \in \mathcal{T}_{0,H}\}$ , and set  $Q_{ad}^H = Q_{ad} \cap Q^H$ . Here we use  $\mathcal{P}_k$  for the space of all polynomials of degree  $\leq k$ . Then we can define the following discrete problem.

**Problem 4.1** Find a function  $p_\varepsilon^H \in Q_{ad}^H$  such that

$$s_\varepsilon(p_\varepsilon^H, q^H - p_\varepsilon^H) \geq l(q^H - p_\varepsilon^H) \quad \forall q^H \in Q_{ad}^H. \quad (4.1)$$

Similar to the continuous case, we have the following well-posedness result for Problem 4.1.

**Proposition 4.2** For each  $H > 0$ , Problem 4.1 has a unique solution  $p_\varepsilon^H$  which depends continuously on all data.

As for error estimate of the finite element solution  $p_\varepsilon^H$  of Problem 4.1, we first derive an abstract error estimate.

**Theorem 4.3** There exists a positive constant  $c$  which is independent of  $\varepsilon > 0$  and  $H > 0$  such that

$$\varepsilon^{1/2} \|p_\varepsilon^H - p_\varepsilon\|_Q \leq c \inf_{q^H \in Q_{ad}^H} (\|q^H - p_\varepsilon\|_Q + \|\mathcal{S}_\varepsilon p_\varepsilon - l\|_Q^{1/2} \|q^H - p_\varepsilon\|_Q^{1/2}). \quad (4.2)$$

**Proof.** By definition,

$$s_\varepsilon(p_\varepsilon^H - p_\varepsilon, p_\varepsilon^H - p_\varepsilon) = \|u_D(p_\varepsilon^H - p_\varepsilon) - u_N(p_\varepsilon^H - p_\varepsilon)\|_{L^2(\Omega)}^2 + \varepsilon \|p_\varepsilon^H - p_\varepsilon\|_Q^2. \quad (4.3)$$

Adding (4.1) and (3.2) with  $q = p_\varepsilon^H$ , we obtain

$$0 \leq s_\varepsilon(p_\varepsilon, p_\varepsilon^H - p_\varepsilon) + s_\varepsilon(p_\varepsilon^H, q^H - p_\varepsilon^H) - l(q^H - p_\varepsilon).$$

Use this inequality,

$$s_\varepsilon(p_\varepsilon^H - p_\varepsilon, p_\varepsilon^H - p_\varepsilon) \leq s_\varepsilon(p_\varepsilon^H, q^H - p_\varepsilon) - l(q^H - p_\varepsilon).$$

Thus,

$$\begin{aligned} & s_\varepsilon(p_\varepsilon^H - p_\varepsilon, p_\varepsilon^H - p_\varepsilon) \\ & \leq s_\varepsilon(p_\varepsilon^H - p_\varepsilon, q^H - p_\varepsilon) + (\mathcal{S}_\varepsilon p_\varepsilon - l, q^H - p_\varepsilon)_Q \\ & \leq \|u_D(p_\varepsilon^H - p_\varepsilon) - u_N(p_\varepsilon^H - p_\varepsilon)\|_{L^2(\Omega)} \|u_D(q^H - p_\varepsilon) - u_N(q^H - p_\varepsilon)\|_{L^2(\Omega)} \\ & \quad + \varepsilon \|p_\varepsilon^H - p_\varepsilon\|_Q \|q^H - p_\varepsilon\|_Q + \|\mathcal{S}_\varepsilon p_\varepsilon - l\|_Q \|q^H - p_\varepsilon\|_Q. \end{aligned}$$

Combining this inequality with (4.3), we deduce that

$$\begin{aligned} & \|u_D(p_\varepsilon^H - p_\varepsilon) - u_N(p_\varepsilon^H - p_\varepsilon)\|_{L^2(\Omega)}^2 + \varepsilon \|p_\varepsilon^H - p_\varepsilon\|_Q^2 \\ & \leq c \left[ \|u_D(q^H - p_\varepsilon) - u_N(q^H - p_\varepsilon)\|_{L^2(\Omega)}^2 + \varepsilon \|q^H - p_\varepsilon\|_Q^2 + \|\mathcal{S}_\varepsilon p_\varepsilon - l\|_Q \|q^H - p_\varepsilon\|_Q \right], \end{aligned}$$

from which we conclude (4.2). ■

A direct consequence from Theorem 4.3 is in the following.

**Corollary 4.4**  $p_\varepsilon^H \rightarrow p_\varepsilon$  in  $Q$  when  $H \rightarrow 0$ ; moreover, if  $p_\varepsilon \in H^1(\Omega_0)$ , we have

$$\varepsilon^{1/2} \|p_\varepsilon^H - p_\varepsilon\|_Q \leq c H^{1/2} \quad (4.4)$$

with  $c$  depending on  $|p_\varepsilon|_{H^1(\Omega_0)}$  but independent of  $\varepsilon > 0$  and  $H > 0$ .

We note that the convergence order in (4.4) is not optimal and an improvement is possible. In fact, using the technique in [18, Lemma 4.7], we have

**Proposition 4.5** *There is a constant  $c$  independent of  $H > 0$  such that*

$$\varepsilon^{1/2} \|p_\varepsilon^H - p_\varepsilon\|_Q \leq c H^{1/2} \|\Pi^H p_\varepsilon - p_\varepsilon\|_Q^{1/2}.$$

Consequently, if  $p_\varepsilon \in H^1(\Omega_0)$ ,

$$\varepsilon^{1/2} \|p_\varepsilon^H - p_\varepsilon\|_Q \leq c H |p_\varepsilon|_{H^1(\Omega_0)}^{1/2}.$$

An examination of the definitions of  $s_\varepsilon(p^H, q)$  and  $l(q)$  shows that we need further to approximate such terms like  $u_D(q, g_1)$ ,  $u_D(q)$ , and  $\tilde{u}_D(g_1)$ . Continuous piecewise linear functions will be utilized for this purpose. Let  $\{\mathcal{T}_h\}_h$  be a regular family of triangulations over domains  $\bar{\Omega} \subset \mathbb{R}^d$  with a meshsize  $h > 0$ . For each triangulation  $\mathcal{T}_h = \{K_h\}$ , define finite element spaces  $V^h$  and  $V_0^h$  as follows.

$$V^h \triangleq \{v \in C(\bar{\Omega}) \mid v|_{K_h} \in \mathcal{P}_1, \forall K_h \in \mathcal{T}_h\}, \quad V_0^h = V^h \cap V_0.$$

For simplicity, in the following discussion, we further assume (2.11). We will use the same symbol  $g_1 \in H^2(\Omega)$  for its trace  $g_1 \in H^{3/2}(\Gamma)$ . Denote by  $\Pi_{V^h} v$  for the piecewise linear interpolant of  $v \in H^2(\Omega)$  and let  $g_1^h = \Pi_{V^h} g_1 \in V^h$ . Moreover, we will use the symbol  $g_1^h + V_0^h$  for the set

$$\{v \in V^h \mid v(a_i) = g_1(a_i) \forall \text{ vertex } a_i \in K_h \cap \Gamma, \forall K_h \in \mathcal{T}_h\}.$$

For each  $q \in Q$ , denote by  $u_D^h(q, g_1^h) \in g_1^h + V_0^h$  the unique solution of

$$a(u, v) = (q, v)_Q \quad \forall v \in V_0^h \quad (4.5)$$

and by  $u_N^h(q, g_2) \in V^h$  the unique solution of

$$a(u, v) = (q, v)_Q + (g_2, v)_{L^2(\Gamma)} \quad \forall v \in V^h, \quad (4.6)$$

where  $a(\cdot, \cdot)$  is defined in (2.4). Similar to the continuous case, we use the symbols  $u_D^h(q)$ ,  $\tilde{u}_D^h(g_1^h)$ ,  $u_N^h(q)$  and  $\tilde{u}_N^h(g_2)$  for  $u_D^h(q, 0)$ ,  $\tilde{u}_D^h(0, g_1^h)$ ,  $u_N^h(q, 0)$  and  $u_N^h(0, g_2)$ , respectively.

Now we give a discrete counterpart of the bilinear form (3.1) and the linear function (2.9). Given  $\varepsilon \geq 0$ , for any  $p, q \in Q$ ,

$$s_\varepsilon^h(p, q) = s^h(p, q) + \varepsilon(p, q)_Q = (u_D^h(p) - u_N^h(p), u_D^h(q) - u_N^h(q))_{L^2(\Omega)} + \varepsilon(p, q)_Q, \quad (4.7)$$

$$l^h(q) = (\tilde{u}_N^h(g_2) - \tilde{u}_D^h(g_1^h), u_D^h(q) - u_N^h(q))_{L^2(\Omega)}. \quad (4.8)$$

Then  $s_\varepsilon^h$  is symmetric and coercive for each  $\varepsilon > 0$ , and both  $s_\varepsilon^h$  and  $l^h$  are uniformly bounded with respect to  $h$ .

We now introduce a full discretization for Problem 4.1.

**Problem 4.6** Find  $p_\varepsilon^{h,H} \in Q_{ad}^H$  such that

$$s_\varepsilon^h(p_\varepsilon^{h,H}, q^H - p_\varepsilon^{h,H}) \geq l^h(q^H - p_\varepsilon^{h,H}) \quad \forall q^H \in Q_{ad}^H. \quad (4.9)$$

The following result holds.

**Proposition 4.7** Problem 4.6 admits a unique solution  $p_\varepsilon^{h,H}$  in  $Q_{ad}^H$ , and the solution depends continuously on the data.

The rest of this section is devoted to an error estimation for Problem 4.6. We first present some preliminary results.

From [17, Theorem 2.4.2.5 and Proposition 2.5.2.3], under the assumptions (2.11), we have the regularity properties:

$$\|u_D(q, g_1)\|_{H^2(\Omega)} \leq c(\|q\|_Q + \|g_1\|_{H^{3/2}(\Gamma)}), \quad (4.10)$$

$$\|u_N(q, g_2)\|_{H^2(\Omega)} \leq c(\|q\|_Q + \|g_2\|_{H^{1/2}(\Gamma)}). \quad (4.11)$$

Using these regularity properties together with Aubin-Nitche trick, we have linear finite element error estimates in the following (see [9, Theorem 3.2.5] for detail).

**Lemma 4.8** For any  $q \in Q$ ,  $g_1 \in H^{3/2}(\Gamma)$  and  $g_2 \in H^{1/2}(\Gamma)$ , there exists a constant  $c$  independent of  $q, g_1, g_2$  and  $h$  such that

$$\|u_D^h(q) - u_D(q)\|_{L^2(\Omega)} \leq c h^2 \|q\|_Q,$$

$$\|u_N^h(q) - u_N(q)\|_{L^2(\Omega)} \leq c h^2 \|q\|_Q,$$

$$\|\tilde{u}_D^h(g_1) - \tilde{u}_D^h(g_1^h)\|_{L^2(\Omega)} \leq c h^2 \|g_1\|_{H^{3/2}(\Gamma)},$$

$$\|\tilde{u}_N^h(g_2) - \tilde{u}_N^h(g_2)\|_{L^2(\Omega)} \leq c h^2 \|g_2\|_{H^{1/2}(\Gamma)}.$$



Let  $\Delta s^h(p, q) = s^h(p, q) - s(p, q)$  and  $\Delta l^h(q) = l^h(q) - l(q)$  for  $p, q \in Q$ . Then we have estimates for  $\Delta s^h$  and  $\Delta l^h$  as follows.

**Lemma 4.9** *For any  $p, q \in Q$ , there is a constant  $c$  which is independent of  $p, q, g_1, g_2$  and  $h > 0$  such that*

$$|\Delta s^h(p, q)| \leq c h^2 \|p\|_Q \|q\|_Q, \quad (4.12)$$

$$|\Delta l^h(q)| \leq c h^2 (\|g_1\|_{H^{3/2}(\Gamma)} + \|g_2\|_{H^{1/2}(\Gamma)}) \|q\|_Q. \quad (4.13)$$

**Proof.** We rewrite  $\Delta s^h(p, q)$  in the following way:

$$\begin{aligned} \Delta s^h(p, q) &= (u_D^h(p) - u_N^h(p), u_D^h(q) - u_N^h(q))_{L^2(\Omega)} - (u_D(p) - u_N(p), u_D(q) - u_N(q))_{L^2(\Omega)} \\ &= (u_D^h(p) - u_N^h(p), u_D^h(q) - u_D(q) - (u_N^h(q) - u_N(q)))_{L^2(\Omega)} \\ &\quad + (u_D^h(p) - u_D(p) - (u_N^h(p) - u_N(p)), u_D(q) - u_N(q))_{L^2(\Omega)}. \end{aligned}$$

Then from Lemma 4.8, and together with regularity properties for  $u_D^h(p)$ ,  $u_D(q)$ ,  $u_N^h(p)$  and  $u_N(q)$ , we obtain (4.12).

Similarly, we decompose  $\Delta l^h$  as follows:

$$\begin{aligned} \Delta l^h(q) &= (\tilde{u}_N^h(g_2) - \tilde{u}_D^h(g_1^h), u_D^h(q) - u_N^h(q))_{L^2(\Omega)} - (\tilde{u}_N(g_2) - \tilde{u}_D(g_1), u_D(q) - u_N(q))_{L^2(\Omega)} \\ &= (\tilde{u}_N^h(g_2) - \tilde{u}_D^h(g_1^h), u_D^h(q) - u_D(q) - (u_N^h(q) - u_N(q)))_{L^2(\Omega)} \\ &\quad + (\tilde{u}_N^h(g_2) - \tilde{u}_N(g_2) - (\tilde{u}_D^h(g_1^h) - \tilde{u}_D(g_1)), u_D(q) - u_N(q))_{L^2(\Omega)}. \end{aligned}$$

From the regularity properties of  $\tilde{u}_N^h(g_2)$ ,  $\tilde{u}_D^h(g_1^h)$ ,  $u_D(q)$  and  $u_N(q)$ , by use of Lemma 4.8, we obtain (4.13). ■

We now present an error bound for the finite element solution from Problem 4.6.

**Theorem 4.10** *There exists a constant  $c > 0$ , independent of  $h$ , such that*

$$\varepsilon \|p_\varepsilon^{h,H} - p_\varepsilon^H\|_Q \leq c h^2 (\|p_\varepsilon\|_Q + \|g_1\|_{H^{3/2}(\Gamma)} + \|g_2\|_{H^{1/2}(\Gamma)}). \quad (4.14)$$

**Proof.** From (4.9) with  $q^H = p_\varepsilon^H$ , we have

$$\begin{aligned} \varepsilon \|p_\varepsilon^{h,H} - p_\varepsilon^H\|_Q^2 &\leq s_\varepsilon^h(p_\varepsilon^{h,H} - p_\varepsilon^H, p_\varepsilon^{h,H} - p_\varepsilon^H) \\ &= s_\varepsilon^h(p_\varepsilon^H, p_\varepsilon^H - p_\varepsilon^{h,H}) - s_\varepsilon^h(p_\varepsilon^{h,H}, p_\varepsilon^H - p_\varepsilon^{h,H}) \\ &\leq s_\varepsilon^h(p_\varepsilon^H, p_\varepsilon^H - p_\varepsilon^{h,H}) + l^h(p_\varepsilon^{h,H} - p_\varepsilon^H). \end{aligned} \quad (4.15)$$

Write

$$l^h(p_\varepsilon^{h,H} - p_\varepsilon^H) = \Delta l^h(p_\varepsilon^{h,H} - p_\varepsilon^H) + l(p_\varepsilon^{h,H} - p_\varepsilon^H).$$

From (4.1) with  $q^H = p_\varepsilon^{h,H}$ , and together with (4.15), we obtain

$$\begin{aligned} \varepsilon \|p_\varepsilon^{h,H} - p_\varepsilon^H\|_Q^2 &\leq s_\varepsilon^h(p_\varepsilon^H, p_\varepsilon^H - p_\varepsilon^{h,H}) + \Delta l^h(p_\varepsilon^{h,H} - p_\varepsilon^H) + s_\varepsilon(p_\varepsilon^H, p_\varepsilon^{h,H} - p_\varepsilon^H) \\ &= \Delta s^h(p_\varepsilon^H, p_\varepsilon^H - p_\varepsilon^{h,H}) + \Delta l^h(p_\varepsilon^{h,H} - p_\varepsilon^H). \end{aligned} \quad (4.16)$$

By use of (4.12) and (4.13) for  $\Delta s^h$  and  $\Delta l^h$ , we obtain (4.14) from (4.16). ■

Combining Theorems 4.3 and 4.10, a full error estimate of finite element approximation is as follows.

**Corollary 4.11** *Let  $p_\varepsilon$  and  $p_\varepsilon^{h,H}$  be the unique solutions of Problems 3.2 and 4.9 respectively. Then*

$$p_\varepsilon^{h,H} \rightarrow p_\varepsilon \text{ in } Q \text{ as } h, H \rightarrow 0.$$

*Moreover, if  $p_\varepsilon \in H^1(\Omega_0)$ , then there exists a constant  $c > 0$ , depending on  $\|p_\varepsilon\|_{H^1(\Omega_0)}$  but independent of  $h$  and  $H$ , such that*

$$\varepsilon \|p_\varepsilon^{h,H} - p_\varepsilon\|_Q \leq c H \varepsilon^{1/2} + c h^2 (\|p_\varepsilon\|_Q + \|g_1\|_{H^{3/2}(\Gamma)} + \|g_2\|_{H^{1/2}(\Gamma)}).$$

At last, we comment that when the solution set  $\mathcal{Z}$  is nonempty, the convergence of  $p_\varepsilon^{h,H}$  to  $p_*$  follows from the triangle inequality

$$\|p_\varepsilon^{h,H} - p_*\|_Q \leq \|p_\varepsilon^{h,H} - p_\varepsilon\|_Q + \|p_\varepsilon - p_*\|_Q$$

in conjunction with Theorem 3.6 and Corollary 4.11.

## 5 Numerical simulation

In this section, we present some numerical results based on our new formulation for the BLT problem. First, we introduce an iterative scheme for this formulation. Then we provide a detailed finite element discretization process of the iterative algorithm. Finally, we show numerical results from two examples.

### 5.1 An iterative algorithm for the BLT problem

Let  $\mathcal{S}$ ,  $l$ ,  $\mathcal{S}_\varepsilon$ , and  $u_D(q, g_1)$ ,  $u_D(q)$ ,  $\tilde{u}_D(g_1)$ ,  $u_N(q, g_2)$ ,  $u_N(q)$  and  $\tilde{u}_N(g_2)$  be given in Sections 2 and 3.

Define two operators  $\mathcal{A}_D$  and  $\mathcal{A}_N$  from  $Q$  to  $H^2(\Omega)$  by

$$\mathcal{A}_D q = u_D(q), \quad \mathcal{A}_N q = u_N(q) \quad \forall q \in Q,$$

and view them as two operators from  $Q$  to  $L^2(\Omega)$ . Set  $\mathcal{A} = \mathcal{A}_D - \mathcal{A}_N$  and denote by  $b = \tilde{u}_N(g_2) - \tilde{u}_D(g_1) \in L^2(\Omega)$ . Then for any  $q \in Q$ ,

$$\mathcal{A}q - b = (\mathcal{A}_D - \mathcal{A}_N)q - b = u_D(q, g_1) - u_N(q, g_2).$$

Denote by  $\mathcal{A}_D^*$  and  $\mathcal{A}_N^*$  the adjoint operators of  $\mathcal{A}_D$  and  $\mathcal{A}_N$ :

$$(\mathcal{A}_D^* v, q)_Q = (v, \mathcal{A}_D q)_{L^2(\Omega)}, \quad (\mathcal{A}_N^* v, q)_Q = (v, \mathcal{A}_N q)_{L^2(\Omega)} \quad \forall v \in L^2(\Omega), \quad q \in Q.$$

Then  $\mathcal{A}^* : L^2(\Omega) \rightarrow Q$  is such that  $\mathcal{A}^* = \mathcal{A}_D^* - \mathcal{A}_N^*$ . Consequently, for any  $p, q \in Q$ ,

$$s_\varepsilon(p, q) = (\mathcal{A}p, \mathcal{A}q)_{L^2(\Omega)} + \varepsilon(p, q)_Q = (\mathcal{A}^* \mathcal{A}p, q)_Q + \varepsilon(p, q)_Q = ((\mathcal{A}^* \mathcal{A} + \varepsilon \mathcal{I})p, q)_Q.$$

Therefore,  $\mathcal{S}_\varepsilon = \mathcal{A}^* \mathcal{A} + \varepsilon \mathcal{I}$ . Similarly,  $l = \mathcal{A}^* b$  comes from  $l(q) = (b, \mathcal{A}q)_{L^2(\Omega)} = (\mathcal{A}^* b, q)_Q$ .

For any  $q \in Q$ , denote by  $u_{DN}(q) = u_D(q, g_1) - u_N(q, g_2)$ , and by  $w_D = w_D(u_{DN}(q)) \in V_0$  and  $w_N = w_N(u_{DN}(q)) \in V$  the solutions of the adjoint variational problems

$$a(v, w_D) = (u_{DN}, v)_{L^2(\Omega)} \quad \forall v \in V_0 \quad (5.1)$$

and

$$a(v, w_N) = (u_{DN}, v)_{L^2(\Omega)} \quad \forall v \in V, \quad (5.2)$$

respectively. Then  $w_D(u_{DN}(q))|_{\Omega_0} = \mathcal{A}_D^*(u_{DN}(q))$  and  $w_N(u_{DN}(q))|_{\Omega_0} = \mathcal{A}_N^*(u_{DN}(q))$ . Thus,

$$\mathcal{A}^*(\mathcal{A}q - b) = (\mathcal{A}_D^* - \mathcal{A}_N^*)(u_{DN}(q)) = (w_D(u_{DN}(q)) - w_N(u_{DN}(q)))|_{\Omega_0}.$$

Let  $P_{ad}$  be the projection operator from  $Q$  onto  $Q_{ad}$ . Following [14, Chapter I, Remark 3.3], we consider an iterative scheme for solving (3.2).

**Algorithm 5.1** 1 Choose  $p^0 \in Q_{ad}$ , set  $k = 0$ .

2 For  $k = 0, 1, \dots$ , with  $p^k \in Q_{ad}$  known,

2.1 solve (2.5) and (2.6) to get  $u_D^k = u_D(p^k, g_1)$  and  $u_N^k = u_N(p^k, g_2)$ ;

2.2 compute  $f^k = u_D^k - u_N^k$ ;

2.3 solve (5.1) and (5.2) with  $u_{DN}(q)$  replaced by  $f^k$  to obtain  $w_D^k = w_D(f^k)$  and  $w_N^k = w_N(f^k)$ ;

2.4 compute  $w^k = w_D^k - w_N^k$ ;

2.5  $\tilde{p}^{k+1} = W_\rho(p^k) = p^k - \rho(w^k|_{\Omega_0} + \varepsilon p^k)$ ;

2.6 project  $\tilde{p}^{k+1}$  onto admissible set  $Q_{ad}$ :  $p^{k+1} = P_{ad}\tilde{p}^{k+1}$ .

Note that under our priori smoothness assumptions on the given data,  $w_D^k \in H^2(\Omega)$ ,  $w_N^k \in H^2(\Omega)$ , and thus  $w^k \in H^2(\Omega)$ . Since  $H^2(\Omega) \hookrightarrow C(\overline{\Omega})$ ,  $w^k|_{\Omega_0}$  is well defined. From (3.7) in the proof of [14, Chapter I, Theorem 3.1], for the iterates to converge, we should select those  $\rho$  which guarantees the operator  $\overline{W}_\rho = W_\rho P_{ad}$  over space  $Q$  is a strictly contractive mapping. By the contractivity of the projection operator  $P_{ad}$ , we only need to show  $W_\rho$  is a strictly contractive mapping. From  $(\mathcal{S}_\varepsilon q, q)_Q \geq \varepsilon \|q\|_Q^2$ , we have, for any  $q_1, q_2 \in Q$ ,

$$\begin{aligned} \|W_\rho(q_1) - W_\rho(q_2)\|_Q^2 &= \|q_1 - q_2\|_Q^2 - 2\rho(\mathcal{S}_\varepsilon(q_1 - q_2), q_1 - q_2)_Q + \rho^2 \|\mathcal{S}_\varepsilon(q_1 - q_2)\|_Q^2 \\ &\leq (1 - 2\rho\varepsilon + \rho^2 \|\mathcal{S}_\varepsilon\|^2) \|q_1 - q_2\|_Q^2, \end{aligned} \quad (5.3)$$

where

$$\|\mathcal{S}_\varepsilon\| = \sup_{p \in Q, p \neq 0} \frac{\|\mathcal{S}_\varepsilon p\|_Q}{\|p\|_Q} = \sup_{p \in Q, p \neq 0} \sup_{q \in Q, q \neq 0} \frac{|(\mathcal{S}_\varepsilon p, q)_Q|}{\|p\|_Q \|q\|_Q} = \sup_{p \in Q, p \neq 0} \sup_{q \in Q, q \neq 0} \frac{|s_\varepsilon(p, q)|}{\|p\|_Q \|q\|_Q}.$$

Thus,  $W_\rho$  is a strict contraction mapping if  $0 < \rho < 2\varepsilon/\|\mathcal{S}_\varepsilon\|^2$  with  $\mathcal{S}_\varepsilon = \mathcal{A}^* \mathcal{A} + \varepsilon \mathcal{I}$ . Moreover, the contraction factor  $(1 - 2\rho\varepsilon + \rho^2 \|\mathcal{S}_\varepsilon\|^2)$  in (5.3) attains its minimum at  $\rho = \varepsilon/\|\mathcal{S}_\varepsilon\|^2$ .

Next, we discuss a discrete analogue of Algorithm 5.1 for the discrete Problem 4.6. For convenience, we assume  $\mathcal{T}_{0,H}$  and  $\mathcal{T}_h$  are consistent, i.e., the triangulation  $\mathcal{T}_{0,H}$  is a restriction of the triangulation  $\mathcal{T}_h$  on  $\overline{\Omega}_0$ . Let  $\{T_i\}_{i=1}^{N_t}$  and  $\{T_{i_k}\}_{k=1}^{N_0}$  be the elements in  $\overline{\Omega}$  and  $\overline{\Omega}_0$ , respectively, where  $N_t$  and  $N_0$  denote the number of elements in  $\Omega$  and  $\Omega_0$  respectively. Any  $q^H \in Q^H$  with  $q^H|_{T_{i_k}} = q_k$ ,  $1 \leq k \leq N_0$ , can be written as  $q^H = \sum_{k=1}^{N_0} q_k \chi_k$ , where  $\chi_k$  is the characteristic function of the element  $T_{i_k}$ . Set  $q = (q_1, q_2, \dots, q_{N_0})^t$ , where  $(\cdot)^t$  stands for transposition of  $(\cdot)$ . As a result, we can define an isomorphism  $J_Q : \mathbb{R}^{N_0} \rightarrow Q^H$  through

$$q^H = J_Q q.$$

Let  $n$  be the number of nodes of the triangulation  $\mathcal{T}_h$ , and let  $\varphi_i(x) \in V^h$ ,  $1 \leq i \leq n$ , be the node basis functions of the finite element space  $V^h$  associated with grid nodes  $x_i$ . Then, for the problems (4.5) and (4.6), the solutions  $u_D^h \in g_1^h + V_0^h$  and  $u_N^h \in V^h$  can be expanded by  $u_D^h = \sum_{i=1}^n u_{D,i} \varphi_i$  and  $u_N^h = \sum_{i=1}^n u_{N,i} \varphi_i$ , respectively, where  $u_{D,i} = u_D^h(x_i)$  and  $u_{N,i} = u_N^h(x_i)$ . Let  $I = \{1, 2, \dots, n\}$ ,  $I_b = \{i \in I \mid x_i \in \Gamma\}$ ,  $I_0 = \{1, 2, \dots, N_0\}$ ,  $I_j = \{k \in I_0 \mid x_j \text{ is a vertex of element } T_{i_k}\}$ ,  $j \in I$ . Moreover, define

$$\begin{aligned} A &= (a_{ji}), \quad a_{ji} = \int_{\Omega} D \nabla \varphi_i \nabla \varphi_j \, dx, \quad i, j \in I, \\ M &= (m_{ji}), \quad m_{ji} = \int_{\Omega} \mu_a \varphi_i \varphi_j \, dx, \quad i, j \in I, \\ F &= (f_{jk}), \quad f_{jk} = \begin{cases} \int_{T_{i_k}} \varphi_j \, dx, & k \in I_j, \\ 0, & k \in I_0 \setminus I_j, \end{cases} \quad j \in I, \\ z &= (z_1, z_2, \dots, z_n)^t, \quad z_j = \int_{\Gamma} g_2 \varphi_j \, ds, \\ K &= A + M. \end{aligned}$$

In what follows, we use the same symbol for a finite element function and its vector representation associated with the given finite element basis functions. Then  $u_D^k$  and  $u_N^k$  in Algorithm 5.1 can be calculated by

$$\begin{aligned} K u_D^k &= F p^k, \quad u_{D,i}^k = g_1(x_i), i \in I_b, \quad u_D^k = \sum_{i=1}^n u_{D,i}^k \varphi_i, \\ K u_N^k &= F p^k + z, \quad u_N^k = \sum_{i=1}^n u_{N,i}^k \varphi_i. \end{aligned}$$

Similarly, if we define

$$C = (c_{ji}), \quad c_{ji} = \int_{\Omega} \varphi_i \varphi_j dx, \quad i, j \in I$$

then

$$\begin{aligned} K w_D^k &= C f^k, \quad w_{D,i}^k = 0, \quad i \in I_b, \quad w_D^k = \sum_{i=1}^n w_{D,i}^k \varphi_i, \\ K w_N^k &= C f^k, \quad w_N^k = \sum_{i=1}^n w_{N,i}^k \varphi_i. \end{aligned}$$

As for the realization of Step 2.6 in Algorithm 5.1, let  $Q_{ad} = \{p \in Q \mid p \geq 0 \text{ a.e. in } \Omega_0\}$ . Then  $q^H \in Q_{ad}^H \Leftrightarrow J_Q^{-1} q^H \in \mathbb{R}_+^{N_0} \triangleq \{q \in \mathbb{R}^{N_0} \mid q \geq 0\}$ . Consequently, the projection operator  $P_{ad}$  has the form:  $P_{ad} q^H = J_Q (\max\{J_Q^{-1} q^H, 0\})$  for any  $q^H \in Q^H$ .

The stopping criterion is as follows. Assume the measurements on data are polluted by noise with noise level  $\delta > 0$ . Then the stop criteria is

$$\|p^{k+1} - p^k\|_Q \leq \mu \delta \quad (5.4)$$

for some constant  $\mu > 1$ . The value of  $\mu$  affects the iterative times and the accuracy in the reconstructed solution.

## 5.2 Numerical examples

We reconstruct light source function based on Problem 4.6 by applying Algorithm 5.1. The computational results presented here are performed by using a MATLAB code in a Dell OPTIPLEX GX280 (32-bit-capable 3.00GHz Pentium 4 CPU, 256MB of RAM). In all tests, we reconstruct the source function solution for different arguments including regularization parameter  $\varepsilon$ , meshsize  $h$ , noise level  $\delta$  and parameter  $\mu$ . We note that all these parameters affect the accuracy of the approximate source function and the iterative number in Algorithm 5.1, etc. Many references, e.g. [5, 15, 23], can be consulted for a proper regularization parameter

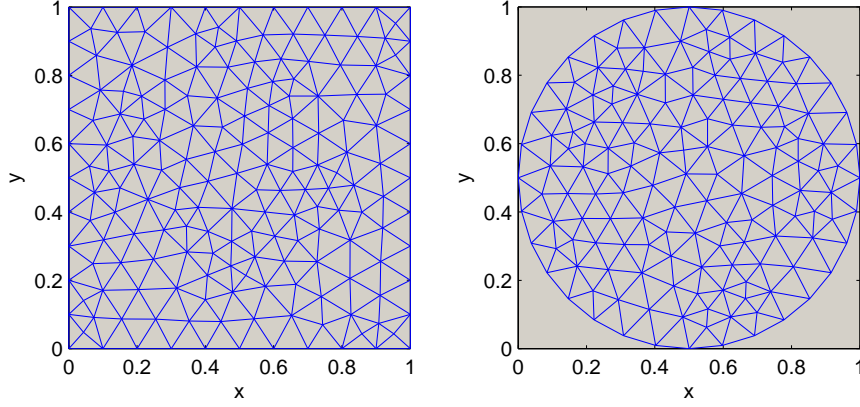


Figure 1: Sample Delaunay triangulations

$\varepsilon$  in the presence of noise. In our examples, we take the regularization parameter  $\varepsilon = 10^{-7}$ . We use

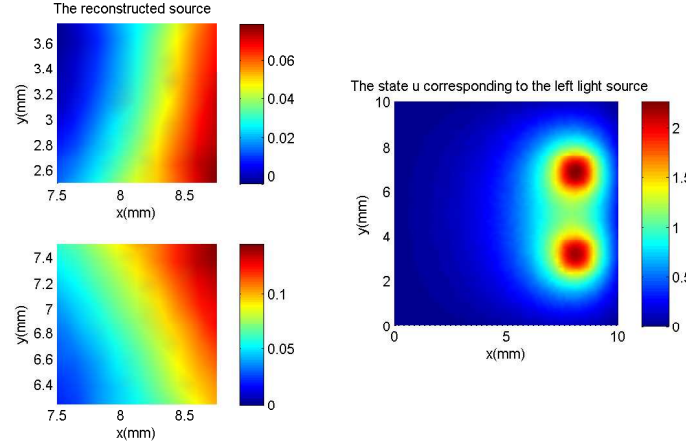
$$u + 2D\partial_\nu u = 0 \quad \text{on } \Gamma, \quad (5.5)$$

as the boundary condition for the PDE (2.1), which is resulted from the condition that the experiments are implemented in a dark environment. Then we take

$$g = -D\partial_\nu u \quad \text{on } \Gamma \quad (5.6)$$

for the measurement on the boundary. From (5.5) and (5.6), we have  $g_1 = 2g$  and  $g_2 = -g$ . Let  $g$ , then  $g_1$  and  $g_2$ , be polluted by noise with level  $\delta = 10\%$ . The admissible set is taken to be  $Q_{ad} = \{q \in Q \mid q \geq 0 \text{ a.e. in } \Omega_0\}$ . We use Delaunay elements for the triangulations  $\{\mathcal{T}_h\}_h$  of the problem domain  $\overline{\Omega}$  and  $\{\mathcal{T}_{0,H}\}_H$  of the permissible domain  $\overline{\Omega}_0$ , and assume they are consistent so that  $H = h$ . See Figure 1 for examples of Delaunay triangulations. Denote by  $p_\varepsilon^{h,h}$  the reconstructed approximate light source and by  $E = \|p_\varepsilon^{h,h} - p\|_Q / \|p\|_Q$  the relative error in  $L^2$ -norm.

In the first example, we let the problem domain  $\Omega = (0, 10) \times (0, 10)$ , and the absorption and reduced scattering coefficients  $\mu_a$  and  $\mu'_s$  be constant in the whole domain with values 0.040 and 1.5 respectively. For our simulation, we take  $p \equiv 1$  pW for the true light source in  $\Omega_* = \{(7.5, 8.75) \times (2.5, 3.75)\} \cup \{(7.5, 8.75) \times (6.25, 7.5)\}$ , and solve the equation (2.1) with boundary condition (5.5) to get the state  $u$  by the FEM in a triangulation with a small meshsize. In this example, we choose  $h = 0.1213$  with 91136 elements and 45889 nodes in  $\overline{\Omega}$  for a triangulation with a small meshsize. Then from (5.5) and (5.6), we set  $g = u/2$  for the measurement  $g$  and thus  $g_1$  and  $g_2$  (with noise) are obtained. Take the permissible domain to be  $\Omega_0 = \Omega_*$ .

Figure 2: Left:  $p_\epsilon^{h,h} - p$  in  $\Omega_0$ ; right:  $u^h$  corresponding to  $p_\epsilon^{h,h}$ Table 1: Iterative number, computation time and relative error E for different  $\rho$  and for meshsize  $h = 0.3750$  in the first example

$\rho$	0.175	0.17	0.15	0.10	0.05	0.01	0.005	0.001
iter-num	161	39	9	4	3	2	1	1
cpu-time	30.61 s	7.95 s	2.19 s	1.05 s	1.03 s	0.84 s	0.66 s	0.66 s
E	55.35%	26.25%	11.88%	9.10%	6.49%	14.08%	18.43%	19.69%

We reconstruct source function solution  $p_\epsilon^{h,h}$  for different parameter  $\rho$  and meshsize  $h$ . Because the norm of the operator  $\mathcal{S}_\epsilon$  is difficult to compute, it is not easy to give an upper bound for  $\rho$ , not to mention the best  $\rho$ . We can take the smallest value  $\rho_{max}$  from those  $\rho$  which make the corresponding sequence  $\{\|p^k - p^{k-1}\|_2\}_k$  nondecreasing during the iteration as an upper bound for  $\rho$ . In this test, a value near 0.05 for  $\rho$  appears to be an advisable choice for a good reconstruction. As for the stopping criterion (5.4), we take  $\mu = 5$ . We plot  $p_\epsilon^{h,h} - p$  and the corresponding state of this reconstructed light source function for  $\rho = 0.05$  and  $h = 0.3750$  with 5696 elements and 2929 nodes in Figure 2. We show the effect of the parameter  $\rho$  on the iterative number, the computation time, and the accuracy of the regularized approximate source in Table 1. The dependence of these terms on the meshsize  $h$  is provided in Table 2.

Table 2: Iterative number, computation time and relative error E for different meshsize and for  $\rho = 0.05$  in the first example

$h$	1.2646	0.6767	0.3750	0.2133
iter-num	2	3	3	4
cpu-time	0.16 s	0.39 s	1.03 s	5.86 s
E	13.63%	10.91%	6.49%	3.89%

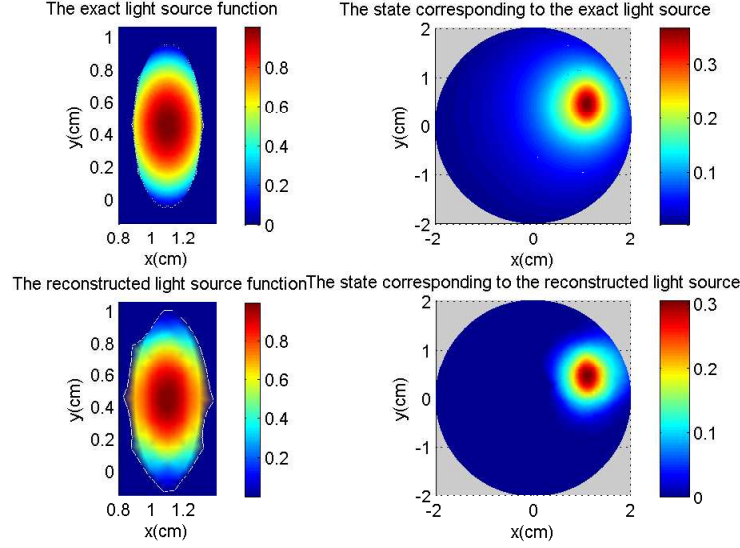


Figure 3: Left:  $p$  and  $p_{\epsilon}^{h,h}$  in  $\Omega_0$ ; right: corresponding state  $u$  and  $u^h$

Table 3: Iterative number, computation time and relative error E for different  $\rho$  and for meshsize  $h = 0.1336$  in the second example

$\rho$	1.1	1.0	0.5	0.1	0.05	0.01	0.001
iter-num	71	16	3	8	8	1	1
cpu-time	12.30 s	3.20 s	1.02 s	1.61 s	1.86 s	0.69 s	0.69 s
E	10.30%	3.51%	0.53%	1.28%	3.27%	7.09%	7.22%

In our second example, let  $\Omega$  be a circle located at origin with radius 2. Let  $\Omega_1 = \{(x, y) \in \Omega \mid \sqrt{x^2 + y^2} < 0.6\}$ ,  $\Omega_2 = \{(x, y) \in \Omega \mid 0.6 < \sqrt{x^2 + y^2} < 2\}$ , and let the absorption and reduced scattering parameters be 0.09 and 2.3 in  $\Omega_1$ , and 0.10 and 1.8 in  $\Omega_2$ . For the measurements  $g_1$  and  $g_2$ , we place a light source with formulation  $p = 1 - 10(x - 1.1)^2 - 4(y - 0.45)^2$  in the domain

$$\Omega_* = \{(x, y) \in \Omega \mid (x - 1.1)^2/0.2^2 + (y - 0.45)^2/0.5^2 = 1\}$$

and take the restriction on the boundary of the corresponding approximate state function obtained by the FEM for meshsize  $h = 0.0431$  with 80384 elements and 40465 nodes as  $2g$ . Again,  $\Omega_0 = \Omega_*$ , and we use  $\mu = 4$  in (5.4) for the stopping criterion. We plot the true light source density distribution  $p$  and an approximate one  $p_{\epsilon}^{h,h}$  as well as their corresponding state in Figure 3 for  $\rho = 0.5$  and  $h = 0.1336$ . Again we show the effect of the parameters  $\rho$  and  $h$  on the iterative number, the time our compute costs and the accuracy of the regularized approximate source in Tables 3 and 4.



Table 4: Iterative number, computation time and relative error E for different meshsize and for  $\rho = 0.50$  in the second example

$h$	0.4667	0.2387	0.1336	0.0758
iter-num	1	2	3	3
cpu-time	0.13 s	0.42 s	1.02 s	4.36 s
E	2.16%	0.86%	0.53%	0.41%

## References

- [1] S. R. Arridge, Optical tomography in medical imaging, *Inverse Problems* **15** (1999), R41–R93.
- [2] K. Atkinson and W. Han, *Theoretical Numerical Analysis: A Functional Analysis Framework*, second edition, Springer-Verlag, New York, Texts in Applied Mathematics, Volume 39, 2005.
- [3] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, third edition, Springer-Verlag, New York, 2008.
- [4] A. J. Chaudhari, et al, Hyperspectral and multispectral bioluminescence optical tomography for small animal imaging, *Phys. Med Biol.* **50** (2005), 5421–5441.
- [5] J. Cheng and M. Yamamoto, One new strategy for a priori choice of regularizing parameters in Tikhonov’s regularization, *Inverse Problems* **16** (2001), 31–38.
- [6] X.-L. Cheng, R.-F. Gong, and W. Han, A generalized mathematical framework for bioluminescence tomography, *Computer Methods in Applied Mechanics and Engineering* **197** (2008), 524–535.
- [7] X.-L. Cheng, R.-F. Gong, and W. Han, Numerical approximation of bioluminescence tomography based on a new formulation, *Journal of Engineering Mathematics* (2008), DOI: 10.1007/s10665-008-9246-y.
- [8] S. R. Cherry, In vivo molecular and genomic imaging: new challenges for imaging physics, *Phys. Med. Biol.* **49** (2004), 13–48.
- [9] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, North-Holland, 1978.
- [10] W.-X. Cong, K. Durairaj, L.-V. Wang, and G. Wang, A Born-type approximation method for bioluminescence tomography, *Med. Phys.* **33** (2006), 679–686.
- [11] A.-X. Cong and G. Wang, Multispectral bioluminescence tomography: methodology and simulation, *Int. J. Biomed. Imag.* **2006**, Article ID 57614, 7 pages.

- [12] W.-X. Cong, et al, A practical reconstruction method for bioluminescence tomography, *Opt. Express* **13** (2005), 6756–6771.
- [13] L. C. Evans, *Partial Differential Equations*, American Mathematical Society, 1998.
- [14] R. Glowinski, *Numerical methods for nonlinear variational problems*, Springer-Verlag, 1983.
- [15] G. H. Golub and U. V. Matt, Tikhonov regularization for large scale problems, in *Scientific Computing*, eds G. H. Golub, S. H. Lui, F. T. Luk and R. J. Plemmons, Springer, 1997, pp. 3–26.
- [16] W. Gong, R. Li, N. Yan and W. Zhao, An improved error analysis for finite element approximation of bioluminescence tomography, *J. Comp. Math.* **26** (2008), 297–309.
- [17] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, 1985.
- [18] W. Han, W.-X. Cong, and G. Wang, Mathematical theory and numerical analysis of bioluminescence tomography, *Inverse Problems* **22** (2006), 1659–1675.
- [19] W. Han, W.-X. Cong, and G. Wang, Mathematical study and numerical simulation of multispectral bioluminescence tomography, *International Journal of Biomedical Imaging* **2006** (2006), Article ID 54390, 10 pages, doi:10.1155/IJBI/2006/54390.
- [20] W. Han, K. Kazmi, W.-X. Cong, and G. Wang, Bioluminescence tomography with optimized optical parameters, *Inverse Problems* **23** (2007), 1215–1228.
- [21] W. Han and G. Wang, Theoretical and numerical analysis on multispectral bioluminescence tomography, *IMA Journal of Applied Mathematics* **72** (2007), 67–85.
- [22] W. Han and G. Wang, Bioluminescence tomography: biomedical background, mathematical theory, and numerical approximation, *Journal of Computational Mathematics* **26** (2008), 324–335.
- [23] P. Hansen, The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.* **14** (1993), 1487–1503.
- [24] A. D. Klose, V. Ntziachristos, and A. H. Hielscher, The inverse source problem based on the radiative transfer equation in optical molecular imaging, *J. Comput. Phys.* **202** (2005), 323–345.
- [25] C. S. Levin, Primer on molecular imaging technology, *European J. Nuclear Med. and Mol. Imag.* **32** (2005), 325–345.
- [26] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer, 1971.

- [27] Y. J. Lv and et al, A multilevel adaptive finite element algorithm for bioluminescence tomography, *Opt. Express* **14** (2006), 8211–8223.
- [28] Y. J. Lv, Spectrally resolved bioluminescence tomography with adaptive finite element analysis: methodology and simulation, *Phys. Med. Biol.* **52** (2007), 4497–4512.
- [29] T. F. Massoud and S. S. Gambhir, Molecular imaging in living subjects: seeing fundamental biological processes in a new light, *Genes Dev.* **17** (2003), 545–580.
- [30] F. Natterer and F. Wübbeling, *Mathematical Methods in Image Reconstruction*, SIAM, Philadelphia, 2001.
- [31] P. Ray, A. M. Wu, and S. S. Gambhir, Optical bioluminescence and positron emission tomography imaging of a novel fusion reporter gene in tumor xenografts of living mice, *Cancer Res.* **63** (2003), 1160–1165.
- [32] T. Troy, D. J. McMullen, L. Sambucetti, and B. Rice, Quantitative comparison of the sensitivity of detection of fluorescent and bioluminescent reporters in animal models. *Mol. Imag.* **3** (2004), 9–23.
- [33] G. Wang, E. A. Hoffman, et al., Development of the first bioluminescent CT scanner, *Radiology* **229(P)** (2003), 566.
- [34] G. Wang, Y. Li, and M. Jiang, Uniqueness theorems in bioluminescence tomography, *Med. Phys.* **31** (2004), 2289–2299.
- [35] Q. Zhang, L. Yin, Y. Tan, Z. Yuan, and H. Jiang, Quantitative bioluminescence tomography guided by diffuse optical tomography, *Opt. Express* **16** (2007), 1481–1486.

# Existence and Uniqueness of the Solution for Degenerate Semilinear Parabolic Equations

W. Y. Chan

*Department of Mathematics, Southeast Missouri State University, Cape Girardeau, MO 63701-6700, USA*

Email address: wchan@semo.edu

## ABSTRACT

For the problem given by  $x^q u_t - (x^\gamma u_x)_x = u^p$  for  $0 < x < a$ ,  $0 < t < T \leq \infty$ ,  $u(x, 0) = u_0(x)$  for  $0 \leq x \leq a$ , and  $u(0, t) = 0 = u(a, t)$  for  $0 < t < T$ , where  $q \geq 0$ ,  $\gamma \in [0, 1)$ ,  $p > 1$ , and  $u_0(x)$  is a nonnegative function for  $0 \leq x \leq a$ , this paper studies existence and uniqueness of the classical solution  $u$  of the problem. Furthermore, the blow-up set of the solution is investigated.

*Key words:* Degenerate parabolic problem; Comparison Theorem; Classical solution; Eigenfunction; Blow-up

## 1. INTRODUCTION

Let  $T \leq \infty$ ,  $q, \gamma, p, a$  be constants such that  $q \geq 0$ ,  $\gamma \in [0, 1)$ ,  $p > 1$ ,  $a > 0$ ,  $D = (0, a)$ ,  $\Omega = D \times (0, T)$ ,  $\bar{D} = [0, a]$ ,  $\bar{\Omega} = \bar{D} \times [0, T)$ ,  $\partial\Omega = (\bar{D} \times \{0\}) \cup (\{0, a\} \times (0, T))$ , and  $Lu = x^q u_t - (x^\gamma u_x)_x$ . The following degenerate semilinear parabolic first initial-boundary value problem is studied,

$$Lu = u^p \text{ in } \Omega, \quad (1)$$

$$u(x, 0) = u_0(x) \text{ on } \bar{D}, \quad u(0, t) = 0 = u(a, t) \text{ for } t \in (0, T), \quad (2)$$

where  $u_0(x)$  is a nonnegative function such that  $u_0(0) = 0 = u_0(a)$  and  $u_0(x) \in C^{2+\alpha}(\bar{D})$  for some  $\alpha \in (0, 1)$ . The study of the problem (1)-(2) is motivated by the research papers of Chen, Liu, and Xie [5], and Floater [7]. Chen, Liu, and Xie studied the blow-up set of the problem (1)-(2) with a nonlocal source term  $\int_0^a u^p dx$ . They showed that  $u$  blows up in a finite time and the blow-up set is  $\bar{D}$ . When  $\gamma = 0$ , Floater studied the blow-up set of  $u$  if  $1 < p \leq q + 1$  and  $u_0(x)$  satisfies the condition  $\frac{d}{dx}(u_0(x)/x) \leq 0$  in  $D$ . He showed that if the solution of the problem (1)-(2) blows up in a finite time, then it blows up only at  $x = 0$ . When  $p > q + 1$ , Chan and Liu [4] proved that  $x = 0$  is not a blow-up point if  $u_0(x)$  satisfies the condition  $u_0'' + u_0^p \geq Ku_0$  in  $D$  for some positive constant  $K$ . In addition, they showed that the blow-up set is a compact subset of  $D$ .

Without the source term  $u^p$ , the problem (1) can be used to illustrate the heat conduction in a rigid slab whose faces at  $x = 0$  and  $x = a$  are in contact with a heat reservoir (cf. Day [6]).  $x^q$  and  $x^\gamma$  are the heat capacity and the thermal conductivity of the slab, respectively.

When  $\gamma = 0$  and  $q = 1$ , the problem (1) can describe the temperature  $u$  of the channel flow of a fluid with a temperature-dependent viscosity in the boundary layer (cf. Chan and Kong [3], and Ockendon [11]); here,  $x$  and  $t$  denote the coordinates perpendicular and parallel to the channel wall, respectively. When  $\gamma = q$ , (1) is transformed into

$$u_t - u_{xx} - \frac{\gamma}{x} u_x = x^{-\gamma} u^p.$$

The behavior of the operator  $\mathfrak{L}u = u_t - u_{xx} - \gamma u_x/x$  was studied by Alexiades [1], and Chan and Chen [2].

In Section 2, we study existence and uniqueness of the classical solution of the problem (1)-(2). In Section 3, we assume that  $u_0(x)$  satisfies the following condition,

$$xu'_0 - u_0 < 0 \text{ in } D. \quad (3)$$

Using an approach different from Floater, we show that  $u$  blows up only at  $x = 0$  when  $1 < p \leq q + 1$ .

## 2. EXISTENCE AND UNIQUENESS OF THE SOLUTION

Firstly, we prove a comparison theorem.

**Lemma 1.** *For any  $\tau \in (0, T)$  and bounded nonnegative function  $B(x, t)$  on  $\bar{D} \times [0, \tau]$ , if  $u$  and  $v \in C(\bar{D} \times [0, \tau]) \cap C^{2,1}(D \times (0, \tau))$ , and*

$$(L - B)u \geq (L - B)v \text{ in } D \times (0, \tau],$$

$$u \geq v \text{ on the parabolic boundary } (\bar{D} \times \{0\}) \cup (\{0, a\} \times (0, \tau]),$$

then  $u \geq v$  on  $\bar{D} \times [0, \tau]$ .

**Proof.** Let  $w = u - v + \varepsilon [1 + x^{(1-\gamma)/2}] e^{ct}$  where  $\varepsilon$  and  $c$  are positive real numbers. Then,  $w > 0$  on  $(\bar{D} \times \{0\}) \cup (\{0, a\} \times (0, \tau])$ . By a direct computation,

$$\begin{aligned} (L - B)w &= (L - B)(u - v) + (L - B)\varepsilon [1 + x^{(1-\gamma)/2}] e^{ct} \\ &\geq \varepsilon e^{ct} \left\{ [1 + x^{(1-\gamma)/2}] (cx^q - B) + \left( \frac{1-\gamma}{2} \right)^2 x^{(-3+\gamma)/2} \right\}. \end{aligned}$$

As  $x \rightarrow 0$ ,  $x^{(-3+\gamma)/2} \rightarrow \infty$ . Let  $k_1 = \max_{(x,t) \in \bar{D} \times [0, \tau]} B$ , and  $s_1$  denote the positive root of

$$\left( \frac{1-\gamma}{2} \right)^2 x^{(-3+\gamma)/2} - [1 + x^{(1-\gamma)/2}] k_1 = 0.$$

Then,

$$(L - B)w > 0 \text{ for } (x, t) \in (0, s_1] \times (0, \tau].$$

If  $s_1 < a$ , we choose

$$c \geq \frac{k_1}{s_1^q}.$$

Therefore,

$$(L - B)w > 0 \text{ in } D \times (0, \tau]. \quad (4)$$

Suppose that  $w \leq 0$  somewhere in  $D \times (0, \tau]$ , then the set

$$\{t \in (0, \tau] : w(x_0, t) \leq 0 \text{ for some } x_0 \in D\}$$

is nonempty. Let  $\underline{t}$  denote its infimum. Since  $w(x, 0) > 0$  on  $\bar{D}$ ,  $0 < \underline{t} \leq \tau$ . Let  $\underline{x}$  denote the smallest  $x \in D$  such that  $w(\underline{x}, \underline{t}) = 0$ . We have  $w_t(\underline{x}, \underline{t}) \leq 0$ . At  $\underline{t}$ ,  $w$  attains its minimum at  $\underline{x}$ , it follows that  $w_x(\underline{x}, \underline{t}) = 0$  and  $w_{xx}(\underline{x}, \underline{t}) \geq 0$ . Therefore, at  $(\underline{x}, \underline{t})$

$$(L - B)w(\underline{x}, \underline{t}) = \underline{x}^q w_t(\underline{x}, \underline{t}) - \underline{x}^\gamma w_{xx}(\underline{x}, \underline{t}) - \gamma \underline{x}^{\gamma-1} w_x(\underline{x}, \underline{t}) - B(\underline{x}, \underline{t}) w(\underline{x}, \underline{t}) \leq 0.$$

It contradicts (4). Hence,  $w > 0$  on  $\bar{D} \times [0, \tau]$ . As  $\varepsilon \rightarrow 0^+$ ,  $u - v \geq 0$  on  $\bar{D} \times [0, \tau]$ .  $\square$

Let  $\theta(x) = x^\nu (a - x)^\nu$  where  $\nu \in (0, 1)$  and  $\nu + \gamma < 1$ ,  $h_0$  be a positive constant such that

$$h_0 \theta(x) \geq u_0(x) \text{ on } \bar{D}, \quad (5)$$

$\tilde{\delta}$  be a positive constant less than  $a/2$  such that there exists some  $t_0$  for which the initial value problem,

$$h'(t) = \frac{(a - \tilde{\delta})^{2\nu p} h^p(t)}{\tilde{\delta}^{q+2\nu}} \text{ for } t \in (0, t_0], \quad h(0) = h_0, \quad (6)$$

has a unique solution, and

$$\zeta - \tilde{\delta}^{\nu p} (a - \tilde{\delta})^{\nu p} h^{p-1}(t) \geq 0 \text{ for } t \in (0, t_0], \quad (7)$$

where

$$\zeta = \min \left\{ \nu(1 - \gamma - \nu) \tilde{\delta}^{\gamma+\nu-2} (a - \tilde{\delta})^\nu, \nu(1 - \nu) (a - \tilde{\delta})^{\gamma+\nu} \tilde{\delta}^{\nu-2} \right\}.$$

Let  $\psi(x, t) = \theta(x) h(t)$ ,  $\omega = D \times (0, t_0]$ ,  $\bar{\omega} = \bar{D} \times [0, t_0]$ , and  $\partial\omega = (\bar{D} \times \{0\}) \cup (\{0, a\} \times (0, t_0])$ . By the construction,  $h'(t) \geq 0$  for  $t \in (0, t_0]$  and  $\psi(x, t) \in C(\bar{\omega}) \cap C^{2,1}(\omega)$ .

**Lemma 2.**  $\psi(x, t) \geq u(x, t)$  on  $\bar{\omega}$ .

**Proof.** By (5),  $\psi(x, 0) \geq u_0(x)$  on  $\bar{D}$ . Also,  $\psi(0, t) = 0$  and  $\psi(a, t) = 0$  for  $t > 0$ . A direct computation gives

$$\begin{aligned} & L\psi - \psi^p \\ &= x^{q+\nu} (a - x)^\nu h'(t) - \nu h(t) \frac{d}{dx} \left[ x^{\gamma+\nu-1} (a - x)^\nu - x^{\gamma+\nu} (a - x)^{\nu-1} \right] \\ &\quad - x^{\nu p} (a - x)^{\nu p} h^p(t) \\ &= x^{q+\nu} (a - x)^\nu h'(t) + \nu(1 - \gamma - \nu) x^{\gamma+\nu-2} (a - x)^\nu h(t) \\ &\quad + \nu(\gamma + 2\nu) x^{\gamma+\nu-1} (a - x)^{\nu-1} h(t) + \nu(1 - \nu) x^{\gamma+\nu} (a - x)^{\nu-2} h(t) \\ &\quad - x^{\nu p} (a - x)^{\nu p} h^p(t). \end{aligned}$$

When  $(x, t) \in (0, \tilde{\delta}] \times (0, t_0]$ , by (7) we have

$$\begin{aligned} L\psi - \psi^p &\geq \nu(1 - \gamma - \nu) \tilde{\delta}^{\gamma+\nu-2} (a - \tilde{\delta})^\nu h(t) - \tilde{\delta}^{\nu p} (a - \tilde{\delta})^{\nu p} h^p(t) \\ &\geq h(t) \left[ \zeta - \tilde{\delta}^{\nu p} (a - \tilde{\delta})^{\nu p} h^{p-1}(t) \right] \\ &\geq 0. \end{aligned}$$

Similarly, when  $(x, t) \in [a - \tilde{\delta}, a) \times (0, t_0]$ ,

$$\begin{aligned} L\psi - \psi^p &\geq \nu(1 - \nu) (a - \tilde{\delta})^{\gamma+\nu} \tilde{\delta}^{\nu-2} h(t) - (a - \tilde{\delta})^{\nu p} \tilde{\delta}^{\nu p} h^p(t) \\ &\geq h(t) \left[ \zeta - \tilde{\delta}^{\nu p} (a - \tilde{\delta})^{\nu p} h^{p-1}(t) \right] \\ &\geq 0. \end{aligned}$$

When  $(x, t) \in (\tilde{\delta}, a - \tilde{\delta}) \times (0, t_0]$ , by (6) we obtain

$$\begin{aligned} L\psi - \psi^p &\geq x^{q+\nu} (a - x)^\nu h'(t) - x^{\nu p} (a - x)^{\nu p} h^p(t) \\ &\geq \tilde{\delta}^{q+2\nu} h'(t) - (a - \tilde{\delta})^{2\nu p} h^p(t) \\ &= 0. \end{aligned}$$

By Lemma 1,  $\psi(x, t) \geq u(x, t)$  on  $\bar{\omega}$ . □

To prove existence and uniqueness of the solution of the problem (1)-(2), let  $\rho(x) \in C^1(\bar{D})$  be a nondecreasing function such that  $\rho(x) = 0$  when  $x \leq 0$  and  $\rho(x) = 1$  when  $x \geq 1$ . Let  $\delta$  be a positive real number such that  $\delta < a/2$ . We also let  $D_\delta = (\delta, a)$ ,  $\omega_\delta = D_\delta \times (0, t_0]$ ,  $\bar{D}_\delta = [\delta, a]$ , and  $\bar{\omega}_\delta = \bar{D}_\delta \times [0, t_0]$ . In addition, let  $\partial\omega_\delta = (\bar{D}_\delta \times \{0\}) \cup (\{\delta, a\} \times (0, t_0])$ ,

$$\rho_\delta(x) = \begin{cases} 0, & x \leq \delta, \\ \rho\left(\frac{x}{\delta} - 1\right), & \delta < x < 2\delta, \\ 1, & x \geq 2\delta, \end{cases}$$

and

$$u_{0\delta}(x) = \rho_\delta(x) u_0(x).$$

We note that

$$\frac{\partial}{\partial \delta} u_{0\delta}(x) = \begin{cases} 0, & x \leq \delta, \\ -\frac{x}{\delta^2} \rho'\left(\frac{x}{\delta} - 1\right) u_0(x), & \delta < x < 2\delta, \\ 0, & x \geq 2\delta. \end{cases}$$

Since  $\rho(x)$  is nondecreasing, it follows that

$$\frac{\partial}{\partial \delta} u_{0\delta}(x) \leq 0. \tag{8}$$

As  $0 \leq \rho_\delta(x) \leq 1$ , we have

$$u_0(x) \geq u_{0\delta}(x) \text{ for } x \in \bar{\omega}_\delta.$$

Now, let us consider the following initial-boundary value problem,

$$Lu_\delta = u_\delta^p \text{ in } \omega_\delta, \quad (9)$$

$$u_\delta(x, 0) = u_{0\delta}(x) \text{ on } \bar{D}_\delta, \quad u_\delta(\delta, t) = 0 = u_\delta(a, t) \text{ for } t \in (0, t_0]. \quad (10)$$

We want to show that the problem (9)-(10) has a classical solution  $u_\delta$  converging to a classical solution  $u$  of the problem (1)-(2) when  $\delta$  tends to 0.

**Lemma 3.** *The problem (1)-(2) has a unique nonnegative solution*

$$u \in C(\bar{\omega}) \cap C^{2+\alpha, 1+\alpha/2}((0, a] \times [0, t_0]).$$

**Proof.** As  $\psi(x, t) \geq u_\delta(x, t)$  on  $\partial\omega_\delta$ , by the maximum principle  $\psi(x, t) \geq u_\delta(x, t) \geq 0$  on  $\bar{\omega}_\delta$ . We note that  $x^{-q+\gamma}$  and  $x^{-q+\gamma-1} \in C^{\alpha, \alpha/2}(\bar{\omega}_\delta)$  for some  $\alpha \in (0, 1)$ .  $x^{-q}u_\delta^p \leq \psi^p/\delta^q$  for some  $(x, t, u_\delta) \in \bar{\omega}_\delta \times R$ . It follows from Theorem 4.2.2 of Ladde, Lakshmikantham, and Vatsala [9, p. 143] that the problem (9)-(10) has a unique solution  $u_\delta \in C^{2+\alpha, 1+\alpha/2}(\bar{\omega}_\delta)$ . If  $\delta_1 > \delta_2 > 0$ , it follows from (8)  $u_{\delta_1} < u_{\delta_2}$  in  $\omega_{\delta_1}$ . Therefore,  $\lim_{\delta \rightarrow 0} u_\delta$  exists for all  $(x, t) \in \bar{\omega}$ . Let  $u(x, t) = \lim_{\delta \rightarrow 0} u_\delta(x, t)$  where  $\psi \geq u \geq 0$  on  $\bar{\omega}$ . We want to show that  $u$  is a solution. For any  $(x_1, t_1) \in \omega$ , there exist sets  $\tilde{E} = [\tilde{b}_1, \tilde{b}_2] \times [0, \tilde{t}_1]$  and  $\hat{E} = [\hat{b}_1, \hat{b}_2] \times [0, \hat{t}_1]$  such that  $(x_1, t_1) \in \tilde{E} \subset \hat{E} \subset \bar{\omega}$  (where  $\tilde{b}_1 > \hat{b}_1 > 0$  and  $\tilde{t}_1 \leq \hat{t}_1 \leq t_0$ ). Since  $u_\delta \leq \psi$  in  $\hat{E}$ , we have for any constant  $\tilde{q} > 1$ ,

$$\text{i. } \|u_\delta\|_{L^{\tilde{q}}(\hat{E})} \leq \|\psi\|_{L^{\tilde{q}}(\hat{E})} \leq k_2 \text{ for some positive constant } k_2,$$

$$\text{ii. For } \tau > 0,$$

$$\left\| \gamma x^{-q+\gamma-1} \right\|_{L^{\tilde{q}}([\hat{b}_1, \hat{b}_2] \times (t, t+\tau))} = \frac{\gamma \left[ \hat{b}_2^{\tilde{q}(-q+\gamma-1)+1} - \hat{b}_1^{\tilde{q}(-q+\gamma-1)+1} \right]^{1/\tilde{q}}}{[\tilde{q}(-q+\gamma-1)+1]^{1/\tilde{q}}} \tau^{1/\tilde{q}}$$

tends to 0 when  $\tau$  approaches 0,

$$\text{iii. } \|x^{-q}u_\delta^p\|_{L^{\tilde{q}}(\hat{E})} \leq \hat{b}_1^{-q} \|\psi^p\|_{L^{\tilde{q}}(\hat{E})}.$$

If we choose  $\tilde{q} > 3/(2-\alpha)$ , by Theorem 4.9.1 of Ladyženskaja, Solonnikov, and Ural'ceva [10, pp. 341-342]  $u_\delta \in W_q^{2,1}(\hat{E})$ . By Theorem 2.3.3 there [10, p. 80],  $W_q^{2,1}(\hat{E}) \hookrightarrow H^{\alpha, \alpha/2}(\hat{E})$ . Thus,  $\|u_\delta\|_{H^{\alpha, \alpha/2}(\hat{E})} \leq k_3$



for some positive constant  $k_3$ . Now,

$$\begin{aligned} & \|x^{-q}u_\delta^p\|_{H^{\alpha,\alpha/2}(\tilde{E})} \\ & \leq \hat{b}_1^{-q} \|\psi\|_\infty^p + \sup_{\substack{(x,t) \in \tilde{E} \\ (\tilde{x},t) \in \tilde{E}}} \frac{x^{-q} |u_\delta^p(x,t) - u_\delta^p(\tilde{x},t)|}{|x - \tilde{x}|^\alpha} \\ & + \sup_{\substack{(x,t) \in \tilde{E} \\ (\tilde{x},t) \in \tilde{E}}} \frac{|u_\delta^p(\tilde{x},t)| |x^{-q} - \tilde{x}^{-q}|}{|x - \tilde{x}|^\alpha} + \sup_{\substack{(x,t) \in \tilde{E} \\ (x,\tilde{t}) \in \tilde{E}}} \frac{x^{-q} |u_\delta^p(x,t) - u_\delta^p(x,\tilde{t})|}{|t - \tilde{t}|^{\alpha/2}}. \end{aligned}$$

By the mean value theorem, we have

$$\|x^{-q}u_\delta^p\|_{H^{\alpha,\alpha/2}(\tilde{E})} \leq \hat{b}_1^{-q} \|\psi\|_\infty^p + p\hat{b}_1^{-q} \|\psi\|_\infty^{p-1} \|u_\delta\|_{H^{\alpha,\alpha/2}(\tilde{E})} + \|\psi\|_\infty^p \|x^{-q}\|_{H^{\alpha,\alpha/2}(\tilde{E})} \leq k_4$$

for some positive constant  $k_4$  which is independent of  $\delta$ . In addition,  $\|x^{-q+\gamma}\|_{H^{\alpha,\alpha/2}(\tilde{E})}$  and  $\|\gamma x^{-q+\gamma-1}\|_{H^{\alpha,\alpha/2}(\tilde{E})}$  are bounded. Then, by Theorem 4.10.1 of Ladyženskaja, Solonnikov, and Ural'ceva [10, pp. 351-352], we have

$$\|u_\delta\|_{H^{2+\alpha,1+\alpha/2}(\tilde{E})} \leq k_5$$

for some positive constant  $k_5$  which is independent of  $\delta$ . This implies that  $u_\delta$ ,  $(u_\delta)_t$ ,  $(u_\delta)_x$ , and  $(u_\delta)_{xx}$  are equicontinuous in  $\tilde{E}$ . By the Ascoli-Arzelà theorem,  $\|u\|_{H^{2+\alpha,1+\alpha/2}(\tilde{E})} \leq k_5$ , and the partial derivatives of  $u$  are the limits of the corresponding partial derivatives of  $u_\delta$ . Since  $\psi \geq u \geq 0$  on  $\bar{\omega}$ , by the Sandwich theorem  $u(0,t) = 0 = u(a,t)$  for  $t \in [0, t_0]$ . Thus,  $u \in C(\bar{\omega}) \cap C^{2+\alpha,1+\alpha/2}((0,a] \times [0, t_0])$ . By Lemma 1, there exists a unique nonnegative solution  $u$  to the problem (1)-(2).  $\square$

Let  $T = \sup\{t_0 : \text{the problem (1)-(2) has a unique nonnegative solution } u \in C(\bar{\omega}) \cap C^{2+\alpha,1+\alpha/2}((0,a] \times [0, t_0])\}$ . A proof similar to that of Theorem 2.5 of Floater [7] gives the following theorem.

**Theorem 4.** *The problem (1)-(2) has a unique nonnegative solution*

$$u \in C(\bar{\Omega}) \cap C^{2+\alpha,1+\alpha/2}((0,a] \times [0, T)).$$

If  $T < \infty$ , then  $u$  is unbounded in  $\Omega$ .

### 3. BLOW-UP OF THE SOLUTION

Let  $\phi(x)$  be the fundamental eigenfunction of the following Sturm-Liouville eigenvalue problem,

$$(x^\gamma \phi')' + \lambda x^q \phi = 0 \text{ in } D, \phi(0) = 0 = \phi(a), \quad (11)$$

where  $\lambda$  is its corresponding eigenvalue. From the result of Chen, Liu, and Xie [5],  $\lambda > 0$  and

$$\phi(x) = k_6 x^{\frac{1-\gamma}{2}} J_{\frac{1-\gamma}{q+2-\gamma}} \left( \frac{2\sqrt{\lambda}}{q+2-\gamma} x^{\frac{q+2-\gamma}{2}} \right),$$

where  $J_{(1-\gamma)/(q+2-\gamma)}(z)$  is the Bessel function of the first kind with order  $(1-\gamma)/(q+2-\gamma)$ , and  $\phi(x)$  is positive in  $D$  for some positive constant  $k_6$ . We choose  $k_6$  such that  $\max_{x \in \bar{D}} \phi(x) = 1$ .

Let  $U(t) = \int_0^a x^q \phi u dx$ . We modify the proof of Theorem 8 of Kaplan [8] to obtain the following blow-up result.

**Lemma 5.** *If  $U(0) > (\lambda a^{pq})^{1/(p-1)}$ , then  $u$  blows up in a finite time.*

**Proof.** Multiply  $\phi(x)$  on both sides of (1), we obtain

$$x^q u_t \phi = (x^\gamma u_x)_x \phi + u^p \phi.$$

Integrate the above equation with respect to  $x$  from 0 to  $a$ , and by (2) and (11) we get

$$\begin{aligned} \left( \int_0^a x^q u \phi dx \right)_t &\geq \int_0^a (x^\gamma \phi')' u dx + \int_0^a u^p \left( \frac{x}{a} \right)^q \phi dx \\ &\geq - \int_0^a \lambda x^q \phi u dx + \int_0^a \left( \frac{x^q}{a^q} \phi u \right)^p dx. \end{aligned}$$

By Jensen's inequality,

$$\left( \int_0^a x^q u \phi dx \right)_t \geq -\lambda \int_0^a x^q \phi u dx + \left( \int_0^a \frac{x^q}{a^q} \phi u dx \right)^p.$$

This inequality is equivalent to

$$\frac{d}{dt} (e^{\lambda t} U) \geq \frac{e^{-\lambda(p-1)t}}{a^{pq}} (e^{\lambda t} U)^p.$$

Integrate the above expression from 0 to  $t$ , it yields

$$(e^{\lambda t} U(t))^{-p+1} - U^{-p+1}(0) \leq \frac{1}{\lambda a^{pq}} [e^{\lambda(1-p)t} - 1],$$

which implies

$$U^{1-p}(t) \leq \frac{1}{\lambda a^{pq}} + \left( U^{1-p}(0) - \frac{1}{\lambda a^{pq}} \right) e^{\lambda(p-1)t}.$$

By the assumption,  $U^{1-p}(0) < 1/(\lambda a^{pq})$ . Thus,  $U(t)$  tends to infinity in a finite time which implies that  $u(x, t)$  blows up in a finite time.  $\square$

Let

$$I = x u_x - u + \tilde{\varepsilon} x^r u^m, \tag{12}$$

where  $r$ ,  $m$ , and  $\tilde{\varepsilon}$  are positive real numbers such that:

$$(a) \quad r \geq q + 2 - \gamma,$$

$$(b) \quad 1 < m < p,$$

(c) let  $\tilde{\varepsilon}$  be a positive real number and

$$\tilde{\varepsilon} \leq \min \left\{ \frac{(r-q-2+\gamma)(r+\gamma+m)+r(m-1)}{ma^r(2r-q-2+\gamma)}, \frac{p-m}{ma^{\gamma+r-2}(2r-q-2+\gamma)} \right\},$$

(d) by (3), we also choose  $\tilde{\varepsilon}$  satisfying

$$xu'_0 - u_0 + \tilde{\varepsilon}x^r u_0^m \leq 0 \text{ for } x \in \bar{D}.$$

**Lemma 6.** *If  $p \leq q+1$  and  $u_0(x)$  satisfies (3), then  $I \leq 0$  on  $\bar{\Omega}$ .*

**Proof.** Since  $\lim_{x \rightarrow 0} xu_x = 0$ ,  $\lim_{x \rightarrow 0} I(x, t) = 0$  for  $t \in (0, T)$ . Define  $I(0, t) = \lim_{x \rightarrow 0} I(x, t)$ . Since  $u(x, t) \geq 0$  in  $\Omega$  and  $u(a, t) = 0$ , it follows that  $u_x(a, t) \leq 0$  for  $t \in (0, T)$ . Then,  $I(a, t) \leq 0$  for  $t \in [0, T]$ . By the condition (d),  $I(x, 0) \leq 0$  for  $x \in \bar{D}$ . Differentiate (12) with respect to  $t$ , it gives

$$I_t = xu_{xt} - u_t + \tilde{\varepsilon}mx^r u^{m-1}u_t. \quad (13)$$

Similarly,

$$I_x = xu_{xx} + \tilde{\varepsilon}mx^r u^{m-1}u_x + \tilde{\varepsilon}rx^{r-1}u^m. \quad (14)$$

By (12), this expression is equivalent to

$$I_x = xu_{xx} + \tilde{\varepsilon}mx^{r-1}u^{m-1}I + \tilde{\varepsilon}(m+r)x^{r-1}u^m - \tilde{\varepsilon}^2mx^{2r-1}u^{2m-1}. \quad (15)$$

Differentiate (14) with respect to  $x$ , it yields

$$\left. \begin{aligned} I_{xx} = xu_{xxx} + (1 + \tilde{\varepsilon}mx^r u^{m-1})u_{xx} + 2\tilde{\varepsilon}rmx^{r-1}u^{m-1}u_x \\ + \tilde{\varepsilon}m(m-1)x^r u^{m-2}(u_x)^2 + \tilde{\varepsilon}r(r-1)x^{r-2}u^m. \end{aligned} \right\} \quad (16)$$

According to (13), (14), and (16), we have

$$\begin{aligned} & x^q I_t - (x^\gamma I_x)_x \\ &= x(x^q u_{xt} - x^\gamma u_{xxx}) - x^q u_t + \tilde{\varepsilon}mx^r u^{m-1}x^q u_t - x^\gamma u_{xx} - \tilde{\varepsilon}mx^{r+\gamma}u^{m-1}u_{xx} \\ & - 2\tilde{\varepsilon}rmx^{\gamma+r-1}u^{m-1}u_x - \tilde{\varepsilon}m(m-1)x^{\gamma+r}u^{m-2}(u_x)^2 - \tilde{\varepsilon}r(r-1)x^{\gamma+r-2}u^m - \gamma x^\gamma u_{xx} \\ & - \tilde{\varepsilon}\gamma mx^{\gamma+r-1}u^{m-1}u_x - \tilde{\varepsilon}\gamma rx^{\gamma+r-2}u^m. \end{aligned}$$

Differentiate (1) with respect to  $x$ , we obtain

$$x^q u_{xt} - x^\gamma u_{xxx} = -qx^{q-1}u_t + 2\gamma x^{\gamma-1}u_{xx} + \gamma(\gamma-1)x^{\gamma-2}u_x + pu^{p-1}xu_x.$$

Using this expression and condition (b), we have

$$\begin{aligned} & x^q I_t - (x^\gamma I_x)_x \\ & \leq -(q+1)x^q u_t + \tilde{\varepsilon}mx^r u^{m-1}x^q u_t + (\gamma-1)x^\gamma u_{xx} + \gamma(\gamma-1)x^{\gamma-1}u_x + pu^{p-1}xu_x \\ & - \tilde{\varepsilon}mx^{r+\gamma}u^{m-1}u_{xx} - \tilde{\varepsilon}(2r+\gamma)mx^{\gamma+r-1}u^{m-1}u_x - \tilde{\varepsilon}r(r-1+\gamma)x^{\gamma+r-2}u^m. \end{aligned}$$

From (1),  $x^q u_t = x^\gamma u_{xx} + \gamma x^{\gamma-1} u_x + u^p$ . Then,

$$\begin{aligned} & x^q I_t - (x^\gamma I_x)_x \\ & \leq -(q+1) (x^\gamma u_{xx} + \gamma x^{\gamma-1} u_x + u^p) + \tilde{\varepsilon} m x^r u^{m-1} (x^\gamma u_{xx} + \gamma x^{\gamma-1} u_x + u^p) \\ & \quad + (\gamma-1) x^\gamma u_{xx} + \gamma(\gamma-1) x^{\gamma-1} u_x + p u^{p-1} x u_x - \tilde{\varepsilon} m x^{r+\gamma} u^{m-1} u_{xx} \\ & \quad - \tilde{\varepsilon} (2r+\gamma) m x^{\gamma+r-1} u^{m-1} u_x - \tilde{\varepsilon} r (r-1+\gamma) x^{\gamma+r-2} u^m. \end{aligned}$$

We simplify the above expression, and according to (12), it yields

$$\begin{aligned} & x^q I_t - (x^\gamma I_x)_x \\ & \leq -(q+2-\gamma) x^\gamma u_{xx} - \gamma(q+2-\gamma) x^{\gamma-1} u_x - (q+1) u^p + \tilde{\varepsilon} m x^r u^{m+p-1} \\ & \quad + p u^{p-1} (I + u - \tilde{\varepsilon} x^r u^m) - 2\tilde{\varepsilon} m r x^{\gamma+r-1} u^{m-1} u_x - \tilde{\varepsilon} r (r-1+\gamma) x^{\gamma+r-2} u^m. \end{aligned}$$

By (15) and (12), the above inequality changes to

$$\begin{aligned} & x^q I_t - (x^\gamma I_x)_x \\ & \leq -(q+2-\gamma) x^{\gamma-1} [I_x - \tilde{\varepsilon} m x^{r-1} u^{m-1} I - \tilde{\varepsilon} (m+r) x^{r-1} u^m + \tilde{\varepsilon}^2 m x^{2r-1} u^{2m-1}] \\ & \quad - \gamma(q+2-\gamma) x^{\gamma-2} (I + u - \tilde{\varepsilon} x^r u^m) - (q+1) u^p + \tilde{\varepsilon} m x^r u^{m+p-1} + p u^{p-1} I + p u^p \\ & \quad - \tilde{\varepsilon} p x^r u^{m+p-1} - 2\tilde{\varepsilon} m r x^{\gamma+r-2} u^{m-1} (I + u - \tilde{\varepsilon} x^r u^m) - \tilde{\varepsilon} r (r-1+\gamma) x^{\gamma+r-2} u^m. \end{aligned}$$

By assumption  $p \leq q+1$ , it gives

$$\left. \begin{aligned} & x^q I_t - (x^\gamma I_x)_x + (q+2-\gamma) x^{\gamma-1} I_x \\ & \leq -\tilde{\varepsilon} m (2r+\gamma-q-2) x^{\gamma+r-2} u^{m-1} I - \gamma(q+2-\gamma) x^{\gamma-2} I + p u^{p-1} I \\ & \quad - \tilde{\varepsilon} x^{\gamma+r-2} u^m [(r-q-2+\gamma)(r+\gamma+m) + r(m-1)] \\ & \quad + \tilde{\varepsilon}^2 m x^{\gamma+2r-2} u^{2m-1} (2r-q-2+\gamma) - \tilde{\varepsilon} (p-m) x^r u^{m+p-1}. \end{aligned} \right\} \quad (17)$$

When  $0 \leq u \leq 1$ ,  $-u^{2m-1} \geq -u^m$  for  $m > 1$ . By condition (a), it yields

$$\begin{aligned} & -\tilde{\varepsilon} x^{\gamma+r-2} u^m [(r-q-2+\gamma)(r+\gamma+m) + r(m-1)] \\ & + \tilde{\varepsilon}^2 m x^{\gamma+2r-2} u^{2m-1} (2r-q-2+\gamma) \\ & \leq -\tilde{\varepsilon} x^{\gamma+r-2} u^{2m-1} [(r-q-2+\gamma)(r+\gamma+m) + r(m-1) - \tilde{\varepsilon} a^r m (2r-q-2+\gamma)]. \end{aligned}$$

Choose  $\tilde{\varepsilon}$  such that  $\tilde{\varepsilon} \leq [(r-q-2+\gamma)(r+\gamma+m) + r(m-1)] / [a^r m (2r-q-2+\gamma)]$ , then (17) becomes

$$\left. \begin{aligned} & x^q I_t - (x^\gamma I_x)_x + (q+2-\gamma) x^{\gamma-1} I_x \\ & \leq -\tilde{\varepsilon} m (2r+\gamma-q-2) x^{\gamma+r-2} u^{m-1} I - \gamma(q+2-\gamma) x^{\gamma-2} I + p u^{p-1} I. \end{aligned} \right\} \quad (18)$$

When  $u > 1$ ,  $u^{2m-1} < u^{m+p-1}$  for  $m < p$ . By condition (a), it gives

$$\begin{aligned} & \tilde{\varepsilon}^2 m x^{\gamma+2r-2} u^{2m-1} (2r - q - 2 + \gamma) - \tilde{\varepsilon} (p - m) x^r u^{m+p-1} \\ & \leq \tilde{\varepsilon} x^r u^{m+p-1} [\tilde{\varepsilon} a^{\gamma+r-2} m (2r - q - 2 + \gamma) - (p - m)]. \end{aligned}$$

Choose  $\tilde{\varepsilon}$  such that  $\tilde{\varepsilon} \leq (p - m) / [a^{\gamma+r-2} m (2r - q - 2 + \gamma)]$ , we have

$$\tilde{\varepsilon}^2 m x^{\gamma+2r-2} u^{2m-1} (2r - q - 2 + \gamma) - \tilde{\varepsilon} (p - m) x^r u^{m+p-1} \leq 0.$$

Therefore, (17) reduces to (18). Hence, if  $\tilde{\varepsilon}$  satisfies condition (c), then (18) is true for  $u \geq 0$ .

Let  $J = I - \eta e^{\beta t}$  where  $\eta$  and  $\beta$  are positive real numbers. Then,  $J(x, t) < 0$  on  $\partial\Omega$ ,  $J_t = I_t - \eta \beta e^{\beta t}$ , and  $J_x = I_x$ . From (18), we have

$$\begin{aligned} & x^q (J_t + \eta \beta e^{\beta t}) - (x^\gamma J_x)_x + (q + 2 - \gamma) x^{\gamma-1} J_x \\ & \leq -\gamma (q + 2 - \gamma) x^{\gamma-2} (J + \eta e^{\beta t}) + p u^{p-1} (J + \eta e^{\beta t}) \\ & \quad - \tilde{\varepsilon} m (2r - q - 2 + \gamma) x^{\gamma+r-2} u^{m-1} (J + \eta e^{\beta t}). \end{aligned}$$

For any  $\tau \in (0, T)$ , let  $M = \max_{(x,t) \in \bar{D} \times [0, \tau]} u$ . By condition (a), we obtain

$$\begin{aligned} & x^q J_t - (x^\gamma J_x)_x + (q + 2 - \gamma) x^{\gamma-1} J_x + \gamma (q + 2 - \gamma) x^{\gamma-2} J - p u^{p-1} J \\ & \quad + \tilde{\varepsilon} m (2r - q - 2 + \gamma) x^{\gamma+r-2} u^{m-1} J \\ & \leq \eta e^{\beta t} [-\beta x^q - \gamma (q + 2 - \gamma) x^{\gamma-2} + p M^{p-1}]. \end{aligned}$$

As  $x \rightarrow 0$ ,  $x^{\gamma-2} \rightarrow \infty$ . Let  $s_2$  denote the positive root of

$$-\gamma (q + 2 - \gamma) x^{\gamma-2} + p M^{p-1} = 0.$$

If  $s_2 < a$ , we choose  $\beta$  such that

$$\beta > \frac{p M^{p-1}}{s_2^q}.$$

Therefore, for  $(x, t) \in \Omega$

$$\begin{aligned} 0 & > x^q J_t - (x^\gamma J_x)_x + (q + 2 - \gamma) x^{\gamma-1} J_x + \gamma (q + 2 - \gamma) x^{\gamma-2} J - p u^{p-1} J \\ & \quad + \tilde{\varepsilon} m (2r - q - 2 + \gamma) x^{\gamma+r-2} u^{m-1} J. \end{aligned}$$

Suppose that  $J \geq 0$  somewhere in  $\Omega$ , then the set

$$\{t \in (0, T) : J(x_2, t) \geq 0 \text{ for some } x_2 \in D\}$$

is nonempty. Let  $\check{t}$  denote its infimum. Since  $J(x, 0) < 0$  on  $\bar{D}$ ,  $0 < \check{t} < T$ . Let  $x_3$  denote the smallest  $x \in D$  such that  $J(x_3, \check{t}) = 0$ . We have  $J_t(x_3, \check{t}) \geq 0$ . At  $\check{t}$ ,  $J$  attains its maximum at  $x_3$ , it follows that  $J_x(x_3, \check{t}) = 0$  and  $J_{xx}(x_3, \check{t}) \leq 0$ . Therefore, at  $(x_3, \check{t})$

$$\begin{aligned} 0 &> x_3^q J_t(x_3, \check{t}) - x_3^\gamma J_{xx}(x_3, \check{t}) + (q + 2 - 2\gamma) x_3^{\gamma-1} J_x(x_3, \check{t}) \\ &\quad + \gamma(q + 2 - \gamma) x_3^{\gamma-2} J(x_3, \check{t}) - p(u(x_3, \check{t}))^{p-1} J(x_3, \check{t}) \\ &\quad + \tilde{\varepsilon} m(2r - q - 2 + \gamma) x_3^{\gamma+r-2} (u(x_3, \check{t}))^{m-1} J(x_3, \check{t}) \\ &\geq 0. \end{aligned}$$

It leads to a contradiction. Therefore,  $J < 0$  on  $\bar{\Omega}$ . As  $\eta \rightarrow 0^+$ ,  $I \leq 0$  on  $\bar{\Omega}$ .  $\square$

**Theorem 7.** *If  $p \leq q + 1$  and  $u_0(x)$  satisfies (3), then  $x = 0$  is the only blow-up point.*

**Proof.** According to Lemma 6,  $xu_x - u \leq -\tilde{\varepsilon}x^r u^m$  on  $\bar{\Omega}$ . It implies

$$x^{-1}u_x - x^{-2}u \leq -\tilde{\varepsilon}x^{r-2}u^m.$$

It is equivalent to

$$\frac{d}{dx}(x^{-1}u) \leq -\tilde{\varepsilon}x^{r+m-2}(x^{-1}u)^m.$$

Let  $x_4$  be a positive real number in  $(0, a]$ . For  $x \in (0, x_4)$ , we integrate the above expression from  $x$  to  $x_4$ , it gives

$$\frac{(x_4^{-1}u(x_4, t))^{-m+1} - (x^{-1}u(x, t))^{-m+1}}{m-1} \geq \tilde{\varepsilon} \frac{(x_4^{r+m-1} - x^{r+m-1})}{r+m-1}.$$

Suppose that  $u$  blows up at  $x_4$ , then  $u(x_4, t) \rightarrow \infty$  as  $t \rightarrow T^-$ . The left hand side of the above expression tends to a non-positive real number. However, the right hand side is a positive real number. It leads to a contradiction. Hence,  $x = 0$  is the only blow-up point.  $\square$

## References

- [1] V. Alexiades, Generalized axially symmetric heat potentials and singular parabolic initial boundary value problems, *Arch. Rational Mech. Anal.*, 79, 325-350 (1982).
- [2] C. Y. Chan and C. S. Chen, A numerical method for semilinear singular parabolic quenching problems, *Quart. Appl. Math.*, 47, 45-57 (1989).
- [3] C. Y. Chan and P. C. Kong, Channel flow of a viscous fluid in the boundary layer, *Quart. Appl. Math.*, 55, 51-56 (1997).

- [4] C. Y. Chan and H. T. Liu, Global existence of solutions for degenerate semilinear parabolic problems, *Nonlinear Anal.*, 34, 617-628 (1998).
- [5] Y. Chen, Q. Liu and C. Xie, Blow-up for degenerate parabolic equations with nonlocal source, *Proc. Amer. Math. Soc.*, 132, 135-145 (2004).
- [6] W. A. Day, Parabolic equations and thermodynamics, *Quart. Appl. Math.*, 50, 523-533 (1992).
- [7] M. S. Floater, Blow-up at the boundary for degenerate semilinear parabolic equations, *Arch. Rational Mech. Anal.*, 114, 57-77 (1991).
- [8] S. Kaplan, On the growth of solutions of quasi-linear parabolic equations, *Commun. Pure Appl. Math.*, 16, 305-330 (1963).
- [9] G. S. Ladde, V. Lakshmikantham and A. S. Vatsala, *Monotone Iterative Techniques for Nonlinear Differential Equations*, Pitman, Boston, Massachusetts, 1985, p. 143.
- [10] O. A. Ladyženskaja, V. A. Solonnikov and N. N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, Amer. Math. Soc., Providence, Rhode Island, 1968, pp. 80, 341-342, and 351-352.
- [11] H. Ockendon, Channel flow with temperature-dependent viscosity and internal viscous dissipation, *J. Fluid Mech.*, 93, 737-746 (1979).

# Generalized Shannon Sampling Method reduces the Gibbs Overshoot in the Approximation of a Step Function

Yufang Hao<sup>†</sup>, Achim Kempf<sup>†‡</sup>

<sup>†</sup>Department of Applied Mathematics, University of Waterloo,  
Waterloo, Ontario, N2L 3G1, Canada

<sup>‡</sup>Department of Physics, University of Queensland,  
St Lucia, Queensland, 4072, Australia

yhao@math.uwaterloo.ca

## Abstract

As is well-known, the Gibbs' overshoot in approximating a step function is irreducible when using conventional Shannon sampling. Here, we consider a generalization of Shannon sampling which allows samples to be taken on non-equidistant points, adapted to the behavior of the function. We show, numerically, that the new method allows one to reduce the Gibbs' overshoot. In a concrete example, the amplitude of the overshoot is reduced by 70%. We study the underlying mathematical structure with a view to eventually determining the ultimate bound on how far the Gibbs' overshoot can be reduced.

**Key Words:** Sampling, Shannon, Gibbs, Self-Adjoint Operators, Step Function.

## 1 Introduction

The Shannon sampling theorem was introduced into information theory by Shannon in 1949 [1]. It has since played a crucial role as the link between continuous and



discrete representations of information, finding ubiquitous use in communication engineering and signal processing. Already before Shannon, the theorem was studied by E. Whittaker and J. Whittaker in 1929, and it was also independently studied in the Russian literature by Kotel'nikov in 1933. Hence, it is also called the theorem of Whittaker-Shannon-Kotel'nikov (WSK). For a review, see [2, 3, 4].

The Shannon sampling theorem states that if a function  $\phi(t)$  is  $\Omega$ -bandlimited, i.e.,  $\phi(t)$  has a frequency upper bound  $\Omega$ , then  $\phi(t)$  can be perfectly reconstructed at all time  $t$  from its sample values  $\{\phi(t_n)\}_n$  taken on a set of sampling points  $\{t_n\}_n$  with an equidistant spacing  $t_{n+1} - t_n = 1/(2\Omega)$  via:

$$\phi(t) = \sum_{n=-\infty}^{\infty} G(t, t_n) \phi(t_n) \quad (1)$$

The function  $G(t, t_n)$  is the so-called reconstruction kernel, and in the case of Shannon, it is the shifted sinc function  $\text{sinc}(2\Omega(t - t_n))$  'centered' at  $t_n$ . The frequency upper bound  $\Omega$  is called the bandwidth, and the sampling rate  $1/(2\Omega)$  is referred to as the Nyquist rate.

In addition to its use for the perfect reconstruction of functions in the space of  $\Omega$ -bandlimited functions, the Shannon sampling theorem has also been widely used to approximate non-bandlimited functions. However, in this case, the Gibbs' phenomenon occurs whenever there is a discontinuous jump point leading, for example, to the Gibbs ringing problems in image compression. The clearest example to illustrate this type of overshoot is the step function  $H(t)$ . See Figure 1.

$$H(t) = \begin{cases} 1 & t > 0 \\ 0 & t = 0 \\ -1 & t < 0 \end{cases}$$

In Figure 1, the step function  $H(t)$  is approximated using Shannon sampling, i.e., as a sum of shifted sinc functions. The equidistant sampling points are at integer multiples of the constant spacing  $\Delta s = 1/(2\Omega)$ . Although the sampling density on the right ( $\Delta s = 0.1$ ) is ten times tighter than the one on the left ( $\Delta s = 1.0$ ), the maximum values of both approximating functions are about 1.0664 with an error of 0.001. We used 1000 terms in (1) in both cases. As is well known, the 6.64% difference to the original step function  $H(t)$ , i.e. the Gibbs' overshoot, can not be further reduced when using Shannon sampling, even when increasing the bandwidth.

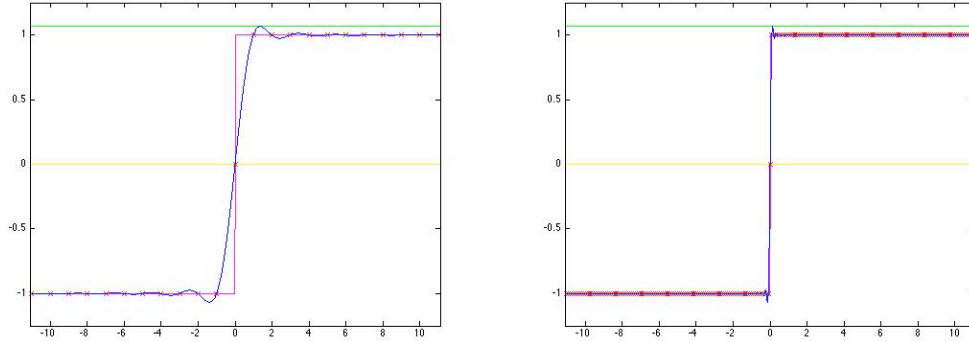


Figure 1: Approximations of the step function by Shannon sampling. The left panel uses a wider sampling spacing of 1.0, while the right panel uses 0.1.

In this paper, we use a generalized sampling theory [5, 6, 7] which allows the sampling and reconstruction of a function on a set of non-equidistant points adapted to the behavior of the function. We show that the new sampling method displays advantages when approximating a step function. See Figure 2.

In Figure 2, we approximated the step function using the generalized sampling method. Outside the interval  $[-10, 10]$ , the sampling points have the same constant spacing  $\Delta s = 1.0$  as the ones on the left in Figure 1. But in the neighborhood interval  $[-10, 10]$  of the jump point, we adjusted the sampling density with 20 extra sampling points. As a result, the maximum value is reduced to 1.0193, which is a 70.9% of reduction in the amplitude of Gibbs' type of overshoot. The amplitude is subject to a numeric error of 0.001, which implies an error of 0.1% in the reduction percentage.

The plot on the right in Figure 2 is a zoom-in of the plot on the left near the jump point. The solid line on the top is the amplitude of the Gibbs overshoot on the uniform lattice in the case of Shannon (which is 1.0664), and the dashed line indicates the amplitude of maximum value with the new generalized sampling theorem (which is 1.0193). This indicates that the new method could be very useful, for example, to reduce Gibbs ringing in image compression. While we have observed the Gibbs overshoot reduction numerically, the analytic reasons and the ultimate limit for the Gibbs overshoot reduction still need to be understood.

In the rest of this paper, we will therefore first recapitulate main features of the

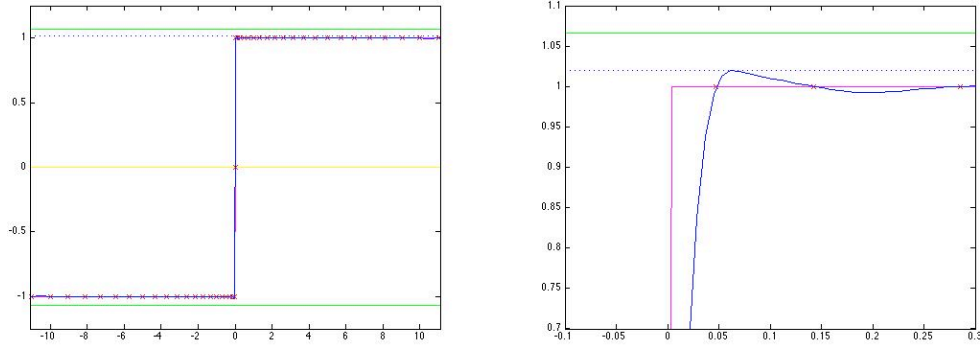


Figure 2: Approximating the step function by the generalized sampling method with non-equidistant sampling points. The right plot zooms in near to the jump point.

Shannon sampling theorem in Section 2, followed by the generalization in Section 3 where we will show that it preserves the key features of Shannon sampling, except for the equidistance restriction on sampling points. In the end, we will show how to choose the non-equidistant sampling lattice adapted to the behaviour of a step function, and to obtain the result in Figure 2 by the generalized sampling method.

## 2 The Shannon Sampling Theorem

In this paper, we call a set of points where we take samples a sampling lattice. The Shannon sampling theorem does not specify where we should start to take samples, but requires that the distance between two adjacent points in one sampling lattice is precisely the constant Nyquist spacing  $1/(2\Omega)$ . Hence, we can parametrize all the sampling lattices as:

$$t_n(\theta) = \frac{n + \theta}{2\Omega}, \quad 0 \leq \theta < 1 \quad (2)$$

Notice that as the parameter  $\theta$  increases from 0 to 1, the sampling points  $\{t_n(\theta)\}$  specified by  $\theta$  increase simultaneously and continuously with the following continuity property:

$$t_n(1) := \lim_{\theta \rightarrow 1^-} t_n(\theta) = t_{n+1}(0) \quad (3)$$

In addition, we can differentiate the sampling points with respect to the parameter  $\theta$  to obtain the velocity with which the sampling points are travelling along the real

line for increasing  $\theta$ . The derivatives obey a similar continuity property:

$$\lim_{\theta \rightarrow 1^-} \frac{dt_n(\theta)}{d\theta} = \left. \frac{dt_{n+1}(\theta)}{d\theta} \right|_{\theta=0} \quad (4)$$

As a result, all the sampling points together,  $\{t_n(\theta)\}_n$  for all integer  $n$  and  $0 \leq \theta < 1$ , cover the real line exactly once.

Of course, given any one-to-one and differentiable function  $\gamma = \gamma(\theta)$  mapping  $[0, 1)$  onto  $[0, 1)$ , we obtain a re-parametrization of the sampling lattices:

$$t_n(\gamma) = \frac{n + \theta(\gamma)}{2\Omega}, \quad 0 \leq \gamma < 1$$

Independent of which parameter we use, the one-parameter family of sampling lattices associated with the Shannon sampling theorem satisfies the properties mentioned above: as the parameter increases, the sampling lattices move to the right along the real line in a continuous manner so that, together, they cover the real line exactly once.

For the particular parameterization with  $\theta$  in (2), the derivatives are precisely the Nyquist sampling rate:

$$\frac{dt_n(\theta)}{d\theta} = \frac{1}{2\Omega}$$

This is not the case for a general parameterization. However, due to the Chain rule, the derivatives at a fixed lattice are always proportional, instead of equal, to the Nyquist sampling rate  $1/(2\Omega)$ :

$$\frac{dt_n(\gamma)}{d\gamma} = \frac{dt_n(\theta)}{d\theta} \frac{d\theta}{d\gamma} = \frac{1}{2\Omega} \frac{d\theta}{d\gamma} \sim \frac{1}{2\Omega}$$

This will be important later.

So far, we saw that the Shannon sampling theorem possesses a natural one-parameter family of Nyquist sampling lattices  $\{t_n(\theta)\}_n$ . Using the reconstruction formula (1), any function in the space of  $\Omega$ -bandlimited functions can be reconstructed from its values taken on each fixed sampling lattice, namely, on  $\{t_n\}_n = \{t_n(\theta)\}_n$  for an arbitrary but fixed  $\theta$ . In other words, treating the reconstruction kernel as a function in  $t$ , the following set of functions

$$g_n^{(\theta)}(t) = G(t, t_n(\theta))$$

forms an orthonormal basis of the function space and hence spans the function space. Most importantly, for each value of the parameter  $\theta$ , from 0 to 1, these bases span

the same function space. As we will see later, this is not trivial, but achievable when we generalize to non-equidistant sampling lattices.

Further, on one fixed lattice, the basis functions interpolate all the points in that lattice:  $g_n^{(\theta)}(t_m(\theta)) = \delta_{nm}$ . More importantly, concerning of stability of reconstruction, notice that the maximum value (or the maximum of the absolute value) of these basis function is always 1 at the point about it is ‘centered’.

### 3 The Generalized Sampling Method

The generalized Shannon sampling theory [5, 6, 7] preserves the key features mentioned in the preceding section, including a one-parameter family of sampling lattices (with non-equidistant sampling points in general), which covers the real line exactly once, the same form of reconstruction formula as in (1), and the corresponding reconstruction kernel (or sets of basis functions) with similar properties as the sinc kernel (or sinc functions).

This generalized sampling theory was initially motivated by the study of the physics of space-time in quantum gravity [8]. The underlying mathematics is based on the functional analytical theory of self-adjoint extensions of unbounded symmetric operators, which will be briefly mentioned at the end of this section.

While generalized Shannon sampling is optimized for non-uniformly spaced sampling points, it is significantly different from conventional non-uniform Shannon sampling [9]. Conventional non-uniform sampling still works within same bandlimited function space as in the case of Shannon, and merely seeks to reconstruct a function in that space from samples taken on a set of non-uniform points. In this case, reconstruction from the set of uniform sampling points is optimally stable among all lattices with the same average sample density. The stability of conventional sampling from non-uniform points can be greatly reduced, as is illustrated by the case of the so-called superoscillatory functions [10]. In approximating a step function, since the conventional Shannon non-uniform sampling uses the same bandlimited function space, the Gibbs’ overshoot can not be made smaller even when using non-uniformly-spaced sampling points.

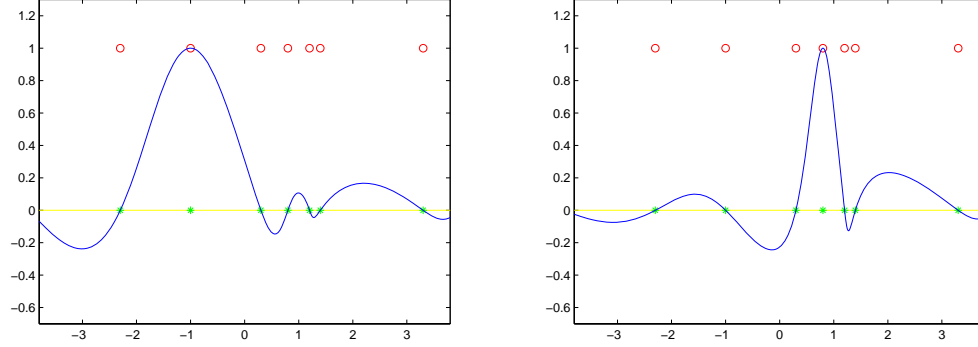


Figure 3: Examples of generalized sinc functions (or reconstruction kernel) on a random non-equidistant sampling lattice

### 3.1 The Generalized Reconstruction Kernel

In the generalized sampling theory, we still deal with one-parameter families of sampling lattices, say  $\{t_n(\alpha)\}_n$ . Now, however, they can possess non-equidistant sampling points. We use the parameter  $\alpha$  to distinguish from the parameter  $\theta$  used in the Shannon sampling theory. Together, the sampling lattices still cover the real line exactly once, and they are differentiable with respect to the parameter  $\alpha$ . We denote the derivative at a point  $t = t_n(\alpha)$  by:

$$t'_n(\alpha) = \frac{dt_n(\alpha)}{d\alpha}$$

The family of sampling lattices still satisfies the continuity conditions (3) and (4). On each fixed lattice, i.e. for a fixed  $\alpha$ , let  $t_n = t_n(\alpha)$ ,  $t'_n = t'_n(\alpha)$ . Then the reconstruction kernel on this lattice was found to read:

$$G(t, t_n) = (-1)^{z(t, t_n)} \frac{\sqrt{t'_n}}{|t - t_n|} \left( \sum_m \frac{t'_m}{(t - t_m)^2} \right)^{-1/2} \quad (5)$$

where  $z(t, t_n(\alpha))$  is the number of the sampling points  $\{t_m(\alpha)\}_m$  between  $t$  and  $t_n(\alpha)$  exclusively.

As a function of  $t$ , for each fixed  $\alpha$ , the set of basis functions

$$\left\{ g_n^{(\alpha)}(t) = G(t, t_n(\alpha)) \right\}_n$$

spans the same function space. Of course, it is generally no longer the space of bandlimited functions. These continuous functions have analogous properties to those

of sinc function: they interpolate all the points in the lattice specified by  $\alpha$

$$g_n^{(\alpha)}(t_m(\alpha)) = G(t_m(\alpha), t_n(\alpha)) = \delta_{mn}$$

and their maximum values are all 1 at the sampling points about which they ‘centered’. Hence, we refer to these basis functions as generalized sinc functions. See Figure 3 for examples.

The fact that each set of basis functions specified by a fixed  $\alpha$  spans the same function space is a remarkable property since, as Figure 3 shows, the shapes of these generalized sinc functions are quite non-trivial.

Let  $\{t_n(\gamma) = t_n(\alpha(\gamma))\}_n$  be a re-parametrization of  $\{t_n(\alpha)\}_n$  where  $\alpha = \alpha(\gamma)$  is a differentiable and strictly increasing function mapping  $[0, 1)$  onto  $[0, 1)$ . For each value of  $\alpha$ , there is one and only one value of  $\gamma$  such that  $t_n(\alpha) = t_n(\gamma)$  for all  $n$ . Hence  $\{t_n(\alpha)\}_n$  and  $\{t_n(\gamma)\}_n$  represent the same sampling lattice, although the derivatives with respect to each parameter are different. However, because we are working on the same family of sampling lattices, independent of the choice of parameterization, we expect the reconstruction kernel on a fix lattice  $\{t_n(\alpha)\}_n = \{t_n(\gamma)\}_n$  to be the same, and hence giving the same function space. This is indeed the case. To see this, use the Chain rule:

$$\frac{dt_n(\gamma)}{d\gamma} = \frac{dt_n(\alpha)}{d\alpha} \frac{d\alpha}{d\gamma}$$

We notice that the derivatives on the fixed lattice respect to different parameters differ only by an  $n$ -independent constant  $C = \frac{d\alpha}{d\gamma}$ . Substitute  $C t'_n(\alpha)$  into (5). The constant  $C$  in the  $\sqrt{t'_n(\alpha)}$ -term in the numerator in front of the infinite sum cancels out the one in  $t'_m(\alpha)$ -term inside the summation since there is a power of  $-1/2$  over the summation.

Therefore, the reconstruction kernel and the set of basis functions on a fixed lattice is invariant under a scalar multiplication of the derivatives, or a re-parametrization of the sampling lattices. As we will see later, this simplifies the work on finding derivatives for a given lattice.

To recover the Shannon sampling theorem as a special case, we choose any uniform sampling lattice  $\{t_n\}_n$  with  $t_{n+1} - t_n = \frac{1}{2\Omega}$  for all  $n$ , together with constant derivatives  $t'_n = C$ . Then the reconstruction kernel in (5) simplifies to the sinc kernel  $\text{sinc}(2\Omega(t - t_n))$ , by using the following trigonometric identity:

$$\frac{\pi^2}{\sin^2(\pi z)} = \sum_{k=-\infty}^{+\infty} \frac{1}{(z - k)^2} \quad (6)$$

### 3.2 One Sampling Lattice and its Derivatives

For practical purposes, one usually only uses one sampling lattice instead of the whole family of lattices, given that one lattice is enough for the reconstruction or interpolation. For example, to approximate a step function, we only need to interpolate samples on one fixed sampling lattice. Hence, it is natural to ask if specifying the whole family of sampling lattice is equivalent to specifying just one, say  $\{t_n(0)\}_n$ . The answer is yes if the derivatives  $\{t'_n(0)\}_n$  on that lattice are also specified.

Indeed without loss of generality, if we are given the sampling lattice and the derivatives at  $\alpha = 0$ , one can generate the whole family of sampling lattices  $\{t_n(\alpha)\}$  by solving for  $t = t_n(\alpha)$  in:

$$\sum_m \frac{t'_m(0)}{t - t_m(0)} = \pi \cot(\pi\alpha) \quad (7)$$

This equation is derived directly from the theory of self-adjoint extensions [5]. Hence it holds for the particular parameterization  $\alpha$  that arises in this way. We will not further discuss the details, but emphasize the fact that finding one sampling lattice and associated derivatives are sufficient to apply the generalized sampling theorem. But which  $\{t_n\}_n$  and  $\{t'_n\}_n$  can arise in the generalized sampling theory?

Arising from the theory of self-adjoint extensions, we have to require the following two restrictions when choosing a sampling lattice  $\{t_n\}_n$ . First, there is a minimum and maximum spacing, namely, there exist positive real numbers  $\delta_{\min}$  and  $\Delta_{\max}$  such that:

$$0 < \delta_{\min} \leq \Delta t_n = t_{n+1} - t_n \leq \Delta_{\max} \quad \text{for all } n \quad (8)$$

In addition, the sampling points  $\{t_n\}_n$  and the corresponding derivatives  $\{t'_n\}_n$  have to satisfy:

$$\sum_n \frac{t'_n}{t_n^2 + 1} = \pi \quad (9)$$

With this initial set of data,  $\{t_n = t_n(0)\}_n$  and  $\{t'_n = t'_n(0)\}_n$ , one can use (7) to generate the whole family of sampling lattices  $\{t_n(\alpha)\}$ , in which each lattice also obeys (8) and (9).

The question now is, once we have a sampling lattice  $\{t_n\}_n$  obeying (8), what would be a suitable choice of the associated derivatives  $t'_n$ ?

First, a derivative  $t'_n(\alpha)$  is the velocity with which the sampling points  $t_n(\alpha)$  are moving to the right along the real line for increasing  $\alpha$  at  $t = t_n(\alpha)$ . Hence a good



candidate for  $t'_n$  is the distance travelled in one period of  $\alpha$ , which is the spacing between two adjacent points. For symmetry, we set  $t'_n$  to be proportional to the average distance between  $t_n$  to its previous and successive points:

$$t'_n \sim \frac{1}{2}(\Delta t_n + \Delta t_{n-1}) = \frac{1}{2}(t_{n+1} - t_{n-1}) \quad (10)$$

The multiplicative constant can be obtained through (9). However, since the reconstruction kernel is independent of a scalar multiplication of the derivatives, we will ignore this scalar prefactor, and use  $t_{n+1} - t_{n-1}$  directly as  $t'_n$  in approximating a function.

### 3.3 Generalizing Principles

The proof of the generalized sampling theorem is outlined in [5, 6]. A more detailed mathematical proof can be found in authors' upcoming paper [7], and the general theory of self-adjoint extensions can be found in [11, 12]. Here, we will outline the idea of this generalization from the functional analytical point of view.

In the Shannon sampling theorem, the set of shifted sinc functions span the space of  $\Omega$ -bandlimited functions. In frequency space, namely, by taking the Fourier transform, the function space corresponds to  $L^2([-\Omega, \Omega])$ , which has a dense subset  $AC^0(-\Omega, \Omega)$ , the set of absolutely continuous functions on  $[-\Omega, \Omega]$  with a vanishing boundary condition. The dense subset  $AC^0(-\Omega, \Omega)$  is an invariant domain under the action of the differential operator  $D := i\frac{d}{d\omega}$ :

$$\Phi(\omega) \in AC^0(-\Omega, \Omega) \implies D\Phi(\omega) = i\Phi'(\omega) \in AC^0(-\Omega, \Omega)$$

The operator  $D$  is a simple symmetric operator with deficiency indices  $(1, 1)$ , and it has therefore a  $U(1)$ -family of self-adjoint extensions. Each self-adjoint extension has a set of discrete eigenvalues  $t_n(\theta) = \frac{n+\theta}{n}$ , which corresponds the set of sampling points. Further, each eigenvalue  $t_n$  has an eigenfunction, which corresponds to the basis function.

The differential operator  $D$  is the multiplicative operator  $T$ ,  $T\phi(t) = t\phi(t)$ , in the time domain, which is also simple symmetric with deficiency indices  $(1, 1)$ . In the case of Shannon, the operator  $T$  is special since all its self-adjoint extensions have equidistant eigenvalues.

In the generalization, we consider a generic symmetric operator  $T$  without equidistant restriction. In functional analytical terms, the sampling theorem expresses the

fact that if a Hilbert space vector is known in the eigenbasis of one self-adjoint extension of  $T$ , then the vector's coefficients in the eigenbases of all other self-adjoint extensions of  $T$  are also determined. Each of those spectra constitutes a set of in general non-equidistant Nyquist sampling points. The sampling kernel then consists of the matrix elements of the unitary transformations which map in between the eigenbases of the self-adjoint extensions of  $T$ , and the function space is the closure of the domain of  $T$ .

## 4 Approximating a Step Function

In summary, to use our generalized sampling theorem to approximate a function, the strategy is:

1. Find a set of points  $\{t_n\}_n$  to form a sampling lattice, depending on the behavior of the function to be approximated. The sampling points have a minimum and maximum spacing (8);
2. Set the corresponding derivatives  $\{t'_n\}_n$  to be  $\frac{1}{2}(t_{n+1} - t_{n-1})$  as in (10);
3. Construct the approximating function by the reconstruction kernel (5) and reconstruction formula (1).

The difficult task is the first step: given a function, how to choose a suitable set of sampling points?

In approximating the step function  $H(t)$ , we have seen that uniform lattices are not optimal, because the step function has a sudden change at the jump point, i.e. a large Gibbs overshoot. Intuitively  $H(t)$  has infinitely large bandwidth at  $t = 0$ , while it has zero bandwidth elsewhere. A windowed Fourier transform shows a regularized behavior of high bandwidth at the step and decreasing bandwidth away from the step. Recall that in the Shannon case, the constant derivative  $t'_n \sim \frac{1}{2\Omega}$  is inverse proportional to the bandwidth  $\Omega$ . Hence, the constant derivatives on a uniform lattice are not matched up with the jump in  $H(t)$ .

When we take samples within the generalized sampling theorem, the derivative, which controls the sampling density, must match the varying bandwidth. Hence, we expect to take

- more samples (at a higher rate) when the function oscillates faster, and

- less samples (at a lower rate) in the period of time with less oscillations.

Thus, for a step function, the spacing between sampling points should be small near the jump point  $t = 0$  where the function has a high bandwidth, and gradually increases to some constant spacing when we sample far away from  $t = 0$ .

The easiest gradually increasing spacing is the linear one. Hence, we will use the sampling lattice with linearly increasing spacing in approximating the step function  $H(t)$  in this paper. We expect this non-equidistant sampling lattice will give a better approximation than the case of Shannon, because now the sampling density matches better with the behavior of the step function  $H(t)$ .

Because  $H(t)$  is anti-symmetric about  $t = 0$ , the sampling points should also be symmetric with respect to  $t = 0$ . Let  $t_0 = 0$  be one of the sampling points, then  $t_{-k} = -t_k$  for all  $k \geq 1$ . From now, we can therefore focus on finding the positive sampling points.

Assume that outside the interval  $[-M\Delta s, M\Delta s]$ , we have equidistant sampling points with a constant spacing  $\Delta s$ :

$$s_m = m\Delta s \quad \forall |m| > M$$

where  $M$  is some positive integer. Inside this interval, we have non-equidistant sampling points with linearly increasing spacing. Assume that we have  $K$  such positive non-equidistant sampling points. Let  $\Delta r$  be the first spacing between  $t_0$  and  $t_1$ , and  $\delta$  be the linear spacing increment. Then

$$\begin{aligned} t_1 &= t_0 + \Delta r = \Delta r \\ t_2 &= t_1 + (\Delta r + \delta) = 2\Delta r + \delta \\ t_3 &= t_2 + (\Delta r + 2\delta) = 3\Delta r + 3\delta \\ &\vdots \\ t_{K-1} &= t_{K-2} + (\Delta r + (K-2)\delta) = (K-1)\Delta r + \frac{1}{2}(K-1)(K-2)\delta \\ t_K &= t_{K-1} + (\Delta r + (K-1)\delta) = K\Delta r + \frac{1}{2}K(K-1)\delta \end{aligned}$$

The chosen sampling lattice is

$$\{s_m\}_{m=-\infty}^{-(M+1)} \cup \{t_k\}_{k=-K}^K \cup \{s_m\}_{m=M+1}^{\infty}$$

The largest non-equidistant sampling point  $t_K$  must match the equidistant sampling point  $s_M = M\Delta s$ , and the spacing must gradually increase to the constant spacing

$\Delta s$ . Hence

$$t_K = K\Delta r + \frac{1}{2}K(K-1)\delta = M\Delta s \quad (11)$$

$$(t_K - t_{K-1}) + \delta = \Delta r + K\delta = \Delta s \quad (12)$$

The difference of 2 times (12) and  $(K-1)$  times (12) gives:

$$(K+1)\Delta r = (2M-K+1)\Delta s \quad (13)$$

Notice that  $\Delta r$  is (and should be much) smaller than  $\Delta s$ . By (13), this means that

$$(K+1) \geq (2M-K+1) \implies M \leq K$$

The smaller  $\Delta r$ , the better approximation. So we want to choose  $K$  as large as possible. Re-write (13) as

$$K = \frac{(2M+1)\Delta s - \Delta r}{\Delta s + \Delta r} = (2M+1) - \frac{2(M+1)}{\Delta s + \Delta r}\Delta r \quad (14)$$

All the terms are positive, so  $K < 2M+1$ , and hence the maximum possible value of  $K$  is  $2M$ . With  $K = 2M$ , we have from (14):

$$\frac{2(M+1)}{\Delta s + \Delta r}\Delta r = 1 \implies \Delta s = (2M+1)\Delta r$$

Substituting this and  $K = 2M$  to (11), we have

$$\Delta r + (2M)\delta = (2M+1)\Delta r \implies \delta = \Delta r$$

In summary, for an optimal result to approximate a step function  $H(t)$  using this linearly increasing spacing method, we pick an interval  $[-M\Delta s, M\Delta s]$  from the uniform lattice with a constant spacing  $\Delta s$ , and replace the uniform sampling points in that interval by a set of points whose adjacent spacing is linearly increasing. There will be twice as many of points as with the uniform lattice, i.e.  $K = 2M$ , and  $\delta = \Delta r = \frac{1}{2M+1}\Delta s$ . In other words, only the  $M$  and  $\Delta s$  are free choices.

In Figure 2, the step function  $H(t)$  is approximated by this method with  $M = 10$  and  $\Delta s = 1.0$ . Comparing this lattice with the uniform lattice with same  $\Delta s$ , we simply replaced the equidistant sampling points on  $[-10, 10]$  by a set of non-equidistant point, whose adjacent spacing linearly decreases toward the jump point  $t = 0$ . As mentioned in the introduction, the maximum value of the new approximation is reduced to 1.0193 with an error of 0.001. This gives about 70 – 71% of reduction in Gibbs' overshoot.

It is important to point out that, using the method of replacing the equidistant points in a finite interval on a uniform lattice with a set of non-equidistant points, the reconstruction kernel can be given in a closed form: using the same trigonometric identity (6) used for recovering the special case of Shannon, we can simplify the infinite sum in (5) into a number subtracting finitely many equidistant points and adding finitely many non-equidistant points. Hence the truncation error only arises from the reconstruction formula in (1), but not from the reconstruction kernel in (5).

## 5 Outlook

Using a non-equidistant lattice with the sampling density linearly increasing toward the jump point, we have seen a significant 70% reduction of the Gibbs' overshoot. Naturally, the question arises whether one can do better with a differently spaced non-equidistant sampling lattice. Is the linear change in sampling density the optimal lattice spacing to match the behavior of a step function? We will address these questions with a more detailed analytical study of the reconstruction kernel.

The best result would be the elimination of all the overshoot. This may not be achievable since the generalized sampling theory considered here assumes that there is no accumulation point of the eigenvalues of these self-adjoint extensions. This corresponds to the assumption of a finite minimum distance in any sampling lattice. However, a further study of the theory of self-adjoint extensions with accumulated eigenvalues may solve the problem.

For practical purposes, the ultimate goal is for any given random function to have an algorithm to determine an optimal sampling lattice, so that the function is (within a pre-specified error) in the function space specified by the lattice.

## Acknowledgment

This work has been supported by NSERC's Discovery, Canada Research Chairs and CGS D2 programs. A.K. and Y.H. gratefully acknowledge the kind hospitality at the University of Queensland, where A.K. is currently on sabbatical.

## References

- [1] C.E. Shannon, Communication in the presence of noise, *Proc. IRE*, 37, 10-21 (1949).
- [2] M. Unser, Sampling - 50 years after Shannon, *Proc. IEEE*, 88, 569-587 (2000).
- [3] A.I. Zayed, *Advances in Shannon's Sampling Theory*, CRC Press, Boca Raton, 1993.
- [4] J.J. Benedetto, *Modern Sampling Theory*, Birkhauser, Boston, 2001.
- [5] Y. Hao, A. Kempf, On a Non-Fourier Generalization of Shannon Sampling Theory, *Proc. of the 10th Canadian Workshop on Information Theory*, 193-196 (2007).
- [6] Y. Hao, A. Kempf, On the Stability of a Generalized Shannon Sampling Theorem, *Proc. of 2008 International Symposium on Information Theorem and its Applications* (2008).
- [7] Y. Hao, A. Kempf, A Generalized Sampling Theorem for Time-Varying Nyquist Rates, in preparation.
- [8] A. Kempf, On fields with finite information density, *Phys. Rev. D*, 69, 124014 (2004).
- [9] J.L. Yen, On non-uniform sampling of bandwidth-limited signals, *IRE Trans. on Circuit Theory*, 251-257 (1956).
- [10] A. Kempf, P.J.S.G. Ferreira, Superoscillations: faster than the Nyquist Rate, *IEEE Trans. Signal Process.*, 54, 3732-3740 (2006).
- [11] N.I. Akhiezer, I.M. Glazman, *Theory of Linear Operators in Hilbert Space*, Dover Publ., New York, 1993.
- [12] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, Toronto, 1978.

---

**Instructions to Contributors**  
**Journal of Concrete and Applicable Mathematics**  
 A quarterly international publication of Eudoxus Press, LLC, of TN.

**Editor in Chief: George Anastassiou**  
 Department of Mathematical Sciences  
 University of Memphis  
 Memphis, TN 38152-3240, U.S.A.

**1. Manuscripts hard copies in triplicate, and in English, should be submitted to the Editor-in-Chief:**

**Prof. George A. Anastassiou**  
 Department of Mathematical Sciences  
 The University of Memphis  
 Memphis, TN 38152, USA.  
 Tel. 901.678.3144  
 e-mail: [ganastss@memphis.edu](mailto:ganastss@memphis.edu)

Authors may want to recommend an associate editor the most related to the submission to possibly handle it.

Also authors may want to submit a list of six possible referees, to be used in case we cannot find related referees by ourselves.

**2. Manuscripts should be typed using any of TEX, LaTeX, AMS-TEX, or AMS-LaTeX and according to EUDOXUS PRESS, LLC. LATEX STYLE FILE. (Click [HERE](#) to save a copy of the style file.) They should be carefully prepared in all respects. Submitted copies should be brightly printed (not dot-matrix), double spaced, in ten point type size, on one side high quality paper 8(1/2)x11 inch. Manuscripts should have generous margins on all sides and should not exceed 24 pages.**

**3. Submission is a representation that the manuscript has not been published previously in this or any other similar form and is not currently under consideration for publication elsewhere. A statement transferring from the authors (or their employers, if they hold the copyright) to Eudoxus Press, LLC, will be required before the manuscript can be accepted for publication. The Editor-in-Chief will supply the necessary forms for this transfer. Such a written transfer of copyright, which previously was assumed to be implicit in the act of submitting a manuscript, is necessary under the U.S. Copyright Law in order for the publisher to carry through the dissemination of research results and reviews as widely and effectively as possible.**

**4. The paper starts with the title of the article, author's name(s) (no titles or degrees), author's affiliation(s) and e-mail addresses. The affiliation should comprise the department, institution (usually university or company), city, state (and/or nation) and mail code.**

**The following items, 5 and 6, should be on page no. 1 of the paper.**

**5. An abstract is to be provided, preferably no longer than 150 words.**

**6. A list of 5 key words is to be provided directly below the abstract. Key words should express the precise content of the manuscript, as they are used for indexing purposes.**

**The main body of the paper should begin on page no. 1, if possible.**

**7. All sections should be numbered with Arabic numerals (such as: 1. INTRODUCTION) .**

**Subsections should be identified with section and subsection numbers (such as 6.1. Second-Value Subheading).**

**If applicable, an independent single-number system (one for each category) should be used to label all theorems, lemmas, propositions, corollaries, definitions, remarks, examples, etc. The label (such as Lemma 7) should be typed with paragraph indentation, followed by a period and the lemma itself.**

**8. Mathematical notation must be typeset. Equations should be numbered consecutively with Arabic numerals in parentheses placed flush right, and should be thusly referred to in the text [such as Eqs.(2) and (5)]. The running title must be placed at the top of even numbered pages and the first author's name, et al., must be placed at the top of the odd numbered pages.**

**9. Illustrations (photographs, drawings, diagrams, and charts) are to be numbered in one consecutive series of Arabic numerals. The captions for illustrations should be typed double space. All illustrations, charts, tables, etc., must be embedded in the body of the manuscript in proper, final, print position. In particular, manuscript, source, and PDF file version must be at camera ready stage for publication or they cannot be considered.**

**Tables are to be numbered (with Roman numerals) and referred to by number in the text. Center the title above the table, and type explanatory footnotes (indicated by superscript lowercase letters) below the table.**

**10. List references alphabetically at the end of the paper and number them consecutively. Each must be cited in the text by the appropriate Arabic numeral in square brackets on the baseline.**

**References should include (in the following order):  
initials of first and middle name, last name of author(s)  
title of article,**



name of publication, volume number, inclusive pages, and year of publication.

Authors should follow these examples:

### **Journal Article**

1. H.H.Gonska, Degree of simultaneous approximation of bivariate functions by Gordon operators, (journal name in italics) *J. Approx. Theory*, 62,170-191(1990).

### **Book**

2. G.G.Lorentz, (title of book in italics) *Bernstein Polynomials* (2nd ed.), Chelsea, New York, 1986.

### **Contribution to a Book**

3. M.K.Khan, Approximation properties of beta operators, in (title of book in italics) *Progress in Approximation Theory* (P.Nevai and A.Pinkus, eds.), Academic Press, New York, 1991, pp.483-495.

11. All acknowledgements (including those for a grant and financial support) should occur in one paragraph that directly precedes the References section.

12. Footnotes should be avoided. When their use is absolutely necessary, footnotes should be numbered consecutively using Arabic numerals and should be typed at the bottom of the page to which they refer. Place a line above the footnote, so that it is set off from the text. Use the appropriate superscript numeral for citation in the text.

13. After each revision is made please again submit three hard copies of the revised manuscript, including in the final one. And after a manuscript has been accepted for publication and with all revisions incorporated, manuscripts, including the TEX/LaTeX source file and the PDF file, are to be submitted to the Editor's Office on a personal-computer disk, 3.5 inch size. Label the disk with clearly written identifying information and properly ship, such as:

Your name, title of article, kind of computer used, kind of software and version number, disk format and files names of article, as well as abbreviated journal name.

Package the disk in a disk mailer or protective cardboard. Make sure contents of disks are identical with the ones of final hard copies submitted!

Note: The Editor's Office cannot accept the disk without the accompanying matching hard copies of manuscript. No e-mail final submissions are allowed! The disk submission must be used.

14. Effective 1 Nov. 2009 for current journal page charges, contact the Editor in Chief. Upon acceptance of the paper an invoice will be sent to the contact author. The fee payment will be due one month from the invoice date. The article will proceed to publication only after the fee is paid. The charges are to be sent, by money order or certified check, in US dollars, payable to Eudoxus Press, LLC, to the address shown on

the Eudoxus [homepage](#).

No galleys will be sent and the contact author will receive one(1) electronic copy of the journal issue in which the article appears.

15. This journal will consider for publication only papers that contain proofs for their listed results.



# **TABLE OF CONTENTS, JOURNAL OF CONCRETE AND APPLICABLE MATHEMATICS, VOL. 8, NO. 3, 2010**

<b>Algorithms for Segmentwise Computation of Forward and Inverse Discrete-time Wavelet Transform, Pavel Rajmic,</b>	<b>393</b>
<b>Approximating common fixed points by an iterative process involving two steps and three mappings, Safeer Hussain Khan,</b>	<b>407</b>
<b>Lawton's conditions on regular low pass filters, A. San Antolin,</b>	<b>416</b>
<b>Strip-saturation Model Solution for Piezoelectric Strip ~ by Quadratically Varying Electric Displacement, R. R. Bhargava, Amit Setia,</b>	<b>426</b>
<b>On the uniqueness of the Fourier projection in <math>L_p</math> spaces, Boris Shekhtman, Leslaw Skrzypek,</b>	<b>439</b>
<b>Object Registration Using Graph Representations of Images, Tamir Nave, Joseph M. Francos, Rami Hagege,</b>	<b>448</b>
<b>Trajectory Tubes of Nonlinear Differential Inclusions and State Estimation Problems, Tatiana F. Filippova,</b>	<b>454</b>
<b>Approximate formulae for fractional derivatives by means of Sinc methods, Tomoaki Okayama, Takayasu Matsuo, Masaaki Sugihara,</b>	<b>470</b>
<b>Boundary type quadrature formulas over axially symmetric regions, Tian-Xiao He,</b>	<b>489</b>
<b>Theoretical Analysis and Numerical Realization of Bioluminescence Tomography, Rongfang Gong, Xiaoliang Cheng, Weimin Han,</b>	<b>504</b>
<b>Existence and Uniqueness of the Solution for Degenerate Semilinear Parabolic Equations, W. Y. Chan,</b>	<b>528</b>
<b>Generalized Shannon Sampling Method reduces the Gibbs Overshoot in the Approximation of a Step Function, Yufang Hao, Achim Kempf,</b>	<b>540</b>

**VOLUME 8, NUMBER 4      OCTOBER 2010**

**ISSN:1548-5390 PRINT,1559-176X ONLINE**



**JOURNAL  
OF CONCRETE  
AND APPLICABLE  
MATHEMATICS**

**EUDOXUS PRESS,LLC**

**SCOPE AND PRICES OF THE JOURNAL**  
**Journal of Concrete and Applicable Mathematics**

A quartely international publication of **Eudoxus Press,LLC**

**Editor in Chief: George Anastassiou**

Department of Mathematical Sciences,  
 University of Memphis  
 Memphis, TN 38152, U.S.A.  
 ganastss@memphis.edu

The main purpose of the "Journal of Concrete and Applicable Mathematics" is to publish high quality original research articles from all subareas of Non-Pure and/or Applicable Mathematics and its many real life applications, as well connections to other areas of Mathematical Sciences, as long as they are presented in a Concrete way. It welcomes also related research survey articles and book reviews. A sample list of connected mathematical areas with this publication includes and is not restricted to: Applied Analysis, Applied Functional Analysis, Probability theory, Stochastic Processes, Approximation Theory, O.D.E, P.D.E, Wavelet, Neural Networks, Difference Equations, Summability, Fractals, Special Functions, Splines, Asymptotic Analysis, Fractional Analysis, Inequalities, Moment Theory, Numerical Functional Analysis, Tomography, Asymptotic Expansions, Fourier Analysis, Applied Harmonic Analysis, Integral Equations, Signal Analysis, Numerical Analysis, Optimization, Operations Research, Linear Programming, Fuzzyness, Mathematical Finance, Stochastic Analysis, Game Theory, Math. Physics aspects, Applied Real and Complex Analysis, Computational Number Theory, Graph Theory, Combinatorics, Computer Science Math. related topics, combinations of the above, etc. In general any kind of Concretely presented Mathematics which is Applicable fits to the scope of this journal. Working Concretely and in Applicable Mathematics has become a main trend in many recent years, so we can understand better and deeper and solve the important problems of our real and scientific world. "Journal of Concrete and Applicable Mathematics" is a peer-reviewed International Quarterly Journal. We are calling for papers for possible publication. The contributor should send three copies of the contribution to the editor in-Chief typed in TEX, LATEX double spaced. [ See: Instructions to Contributors]

**Journal of Concrete and Applicable Mathematics(JCAAM)**

**ISSN:1548-5390 PRINT, 1559-176X ONLINE.**

is published in January, April, July and October of each year by

**EUDOXUS PRESS,LLC,**

1424 Beaver Trail Drive, Cordova, TN38016, USA,

Tel.001-901-751-3553

anastassioug@yahoo.com

<http://www.EudoxusPress.com>.

**Visit also [www.msci.memphis.edu/~ganastss/jcaam](http://www.msci.memphis.edu/~ganastss/jcaam).**

**Webmaster: Ray Clapsadle**

**Annual Subscription Current Prices:** For USA and Canada, Institutional: Print \$400, Electronic \$250, Print and Electronic \$450. Individual: Print \$150, Electronic

\$80,Print &Electronic \$200.For any other part of the world add \$50 more to the above prices for Print.

Single article PDF file for individual \$15.Single issue in PDF form for individual \$60.

No credit card payments.Only certified check,money order or international check in US dollars are acceptable.

Combination orders of any two from JoCAAA,JCAAM,JAFa receive 25% discount,all three receive 30% discount.

**Copyright**©2010 by Eudoxus Press,LLC all rights reserved.JCAAM is printed in USA.

**JCAAM is reviewed and abstracted by AMS Mathematical Reviews,MATHSCI,and Zentralblatt MATH.**

It is strictly prohibited the reproduction and transmission of any part of JCAAM and in any form and by any means without the written permission of the publisher.It is only allowed to educators to Xerox articles for educational purposes.The publisher assumes no responsibility for the content of published papers.

***JCAAM IS A JOURNAL OF RAPID PUBLICATION***

---

## Editorial Board

### Associate Editors

---

**Editor in -Chief:**

George Anastassiou  
 Department of Mathematical Sciences  
 The University Of Memphis  
 Memphis, TN 38152, USA  
 tel. 901-678-3144, fax 901-678-2480  
 e-mail ganastss@memphis.edu  
[www.msci.memphis.edu/~anastasg/anlyjour.htm](http://www.msci.memphis.edu/~anastasg/anlyjour.htm)  
 Areas: Approximation Theory,  
 Probability, Moments, Wavelet,  
 Neural Networks, Inequalities, Fuzzyness.

**Associate Editors:**

1) Ravi Agarwal  
 Florida Institute of Technology  
 Applied Mathematics Program  
 150 W. University Blvd.  
 Melbourne, FL 32901, USA  
[agarwal@fit.edu](mailto:agarwal@fit.edu)  
 Differential Equations, Difference  
 Equations,  
 Inequalities

2) Drumi D. Bainov  
 Medical University of Sofia  
 P.O. Box 45, 1504 Sofia, Bulgaria  
[drumibainov@yahoo.com](mailto:drumibainov@yahoo.com)  
 Differential Equations, Optimal Control,  
 Numerical Analysis, Approximation Theory

3) Carlo Bardaro  
 Dipartimento di Matematica & Informatica  
 Università di Perugia  
 Via Vanvitelli 1  
 06123 Perugia, ITALY  
 tel. +390755855034, +390755853822,  
 fax +390755855024  
[bardaro@unipg.it](mailto:bardaro@unipg.it) ,  
[bardaro@dipmat.unipg.it](mailto:bardaro@dipmat.unipg.it)  
 Functional Analysis and Approximation Th.,  
 Summability, Signal Analysis, Integral  
 Equations,  
 Measure Th., Real Analysis

4) Francoise Bastin  
 Institute of Mathematics  
 University of Liege  
 4000 Liege

21) Gustavo Alberto Perla Menzala  
 National Laboratory of Scientific Computation  
 LNCC/MCT  
 Av. Getulio Vargas 333  
 25651-075 Petropolis, RJ  
 Caixa Postal 95113, Brasil  
 and

Federal University of Rio de Janeiro  
 Institute of Mathematics  
 RJ, P.O. Box 68530 Rio de Janeiro, Brasil  
[perla@lncc.br](mailto:perla@lncc.br) and [perla@im.ufrj.br](mailto:perla@im.ufrj.br)  
 Phone 55-24-22336068, 55-21-25627513 Ext 224  
 FAX 55-24-22315595  
 Hyperbolic and Parabolic Partial Differential  
 Equations,  
 Exact controllability, Nonlinear Lattices and  
 Global  
 Attractors, Smart Materials

22) Ram N. Mohapatra  
 Department of Mathematics  
 University of Central Florida  
 Orlando, FL 32816-1364  
 tel. 407-823-5080  
[ramm@pegasus.cc.ucf.edu](mailto:ramm@pegasus.cc.ucf.edu)  
 Real and Complex analysis, Approximation Th.,  
 Fourier Analysis, Fuzzy Sets and Systems

23) Rainer Nagel  
 Arbeitsbereich Funktionalanalysis  
 Mathematisches Institut  
 Auf der Morgenstelle 10  
 D-72076 Tuebingen  
 Germany  
 tel. 49-7071-2973242  
 fax 49-7071-294322  
[rana@fa.uni-tuebingen.de](mailto:rana@fa.uni-tuebingen.de)  
 evolution equations, semigroups, spectral th.,  
 positivity

24) Panos M. Pardalos  
 Center for Appl. Optimization  
 University of Florida  
 303 Weil Hall  
 P.O. Box 116595  
 Gainesville, FL 32611-6595  
 tel. 352-392-9011  
[pardalos@ufl.edu](mailto:pardalos@ufl.edu)  
 Optimization, Operations Research



## BELGIUM

f.bastin@ulg.ac.be  
Functional Analysis, Wavelets

5) Yeol Je Cho  
Department of Mathematics Education  
College of Education  
Gyeongsang National University  
Chinju 660-701

## KOREA

tel. 055-751-5673 Office,  
055-755-3644 home,  
fax 055-751-6117  
yjcho@nongae.gsnu.ac.kr  
Nonlinear operator Th., Inequalities,  
Geometry of Banach Spaces

6) Sever S. Dragomir  
School of Communications and Informatics  
Victoria University of Technology  
PO Box 14428  
Melbourne City M.C  
Victoria 8001, Australia  
tel 61 3 9688 4437, fax 61 3 9688 4050  
sever.dragomir@vu.edu.au,  
sever@sci.vu.edu.au  
Math. Analysis, Inequalities, Approximation  
Th.,  
Numerical Analysis, Geometry of Banach  
Spaces,  
Information Th. and Coding

7) Angelo Favini  
Università di Bologna  
Dipartimento di Matematica  
Piazza di Porta San Donato 5  
40126 Bologna, ITALY  
tel. ++39 051 2094451  
fax. ++39 051 2094490  
favini@dm.unibo.it  
Partial Differential Equations, Control  
Theory,  
Differential Equations in Banach Spaces

8) Claudio A. Fernandez  
Facultad de Matematicas  
Pontificia Universidad Católica de Chile  
Vicuna Mackenna 4860  
Santiago, Chile  
tel. ++56 2 354 5922  
fax. ++56 2 552 5916  
cfernand@mat.puc.cl  
Partial Differential Equations,  
Mathematical Physics,  
Scattering and Spectral Theory

25) Svetlozar T. Rachev  
Dept. of Statistics and Applied Probability  
Program

University of California, Santa Barbara  
CA 93106-3110, USA  
tel. 805-893-4869  
rachev@pstat.ucsb.edu

## AND

Chair of Econometrics and Statistics  
School of Economics and Business Engineering  
University of Karlsruhe  
Kollegium am Schloss, Bau II, 20.12, R210  
Postfach 6980, D-76128, Karlsruhe, Germany  
tel. 011-49-721-608-7535  
rachev@lsoe.uni-karlsruhe.de  
Mathematical and Empirical Finance,  
Applied Probability, Statistics and Econometrics

26) John Michael Rassias  
University of Athens  
Pedagogical Department  
Section of Mathematics and Informatics  
20, Hippocratous Str., Athens, 106 80, Greece

Address for Correspondence

4, Agamemnonos Str.  
Aghia Paraskevi, Athens, Attikis 15342 Greece  
jrassias@primedu.uoa.gr  
jrassias@tellas.gr  
Approximation Theory, Functional Equations,  
Inequalities, PDE

27) Paolo Emilio Ricci  
Università degli Studi di Roma "La Sapienza"  
Dipartimento di Matematica-Istituto  
"G. Castelnuovo"  
P.le A. Moro, 2-00185 Roma, ITALY  
tel. ++39 0649913201, fax ++39 0644701007  
riccip@uniroma1.it, Paoloemilio.Ricci@uniroma1.it  
Orthogonal Polynomials and Special functions,  
Numerical Analysis, Transforms, Operational  
Calculus,  
Differential and Difference equations

28) Cecil C. Rousseau  
Department of Mathematical Sciences  
The University of Memphis  
Memphis, TN 38152, USA  
tel. 901-678-2490, fax 901-678-2480  
ccrousse@memphis.edu  
Combinatorics, Graph Th.,  
Asymptotic Approximations,  
Applications to Physics

29) Tomasz Rychlik

- 9) A.M.Fink  
Department of Mathematics  
Iowa State University  
Ames, IA 50011-0001, USA  
tel.515-294-8150  
fink@math.iastate.edu  
Inequalities, Ordinary Differential Equations
- 10) Sorin Gal  
Department of Mathematics  
University of Oradea  
Str. Armatei Romane 5  
3700 Oradea, Romania  
galso@uoradea.ro  
Approximation Th., Fuzzyness, Complex Analysis
- 11) Jerome A. Goldstein  
Department of Mathematical Sciences  
The University of Memphis,  
Memphis, TN 38152, USA  
tel.901-678-2484  
jgoldste@memphis.edu  
Partial Differential Equations, Semigroups of Operators
- 12) Heiner H. Gonska  
Department of Mathematics  
University of Duisburg  
Duisburg, D-47048  
Germany  
tel.0049-203-379-3542 office  
gonska@informatik.uni-duisburg.de  
Approximation Th., Computer Aided Geometric Design
- 13) Dmitry Khavinson  
Department of Mathematical Sciences  
University of Arkansas  
Fayetteville, AR 72701, USA  
tel.(479)575-6331, fax(479)575-8630  
dmitry@uark.edu  
Potential Th., Complex Analysis, Holomorphic PDE,  
Approximation Th., Function Th.
- 14) Virginia S. Kiryakova  
Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
virginia@diogenes.bg  
Special Functions, Integral Transforms, Fractional Calculus
- 15) Hans-Bernd Knoop  
Institute of Mathematics  
Polish Academy of Sciences  
Chopina 12, 87100 Torun, Poland  
T.Rychlik@impan.gov.pl  
Mathematical Statistics, Probabilistic Inequalities
- 30) Bl. Sendov  
Institute of Mathematics and Informatics  
Bulgarian Academy of Sciences  
Sofia 1090, Bulgaria  
bsendov@bas.bg  
Approximation Th., Geometry of Polynomials, Image Compression
- 31) Igor Shevchuk  
Faculty of Mathematics and Mechanics  
National Taras Shevchenko  
University of Kyiv  
252017 Kyiv  
UKRAINE  
shevchuk@univ.kiev.ua  
Approximation Theory
- 32) H.M. Srivastava  
Department of Mathematics and Statistics  
University of Victoria  
Victoria, British Columbia V8W 3P4  
Canada  
tel.250-721-7455 office, 250-477-6960 home,  
fax 250-721-8962  
harimsri@math.uvic.ca  
Real and Complex Analysis, Fractional Calculus and Appl.,  
Integral Equations and Transforms, Higher Transcendental Functions and Appl., q-Series and q-Polynomials, Analytic Number Th.
- 33) Stevo Stevic  
Mathematical Institute of the Serbian Acad. of Science  
Knez Mihailova 35/I  
11000 Beograd, Serbia  
sstevic@ptt.yu; sstevo@matf.bg.ac.yu  
Complex Variables, Difference Equations, Approximation Th., Inequalities
- 34) Ferenc Szidarovszky  
Dept. Systems and Industrial Engineering  
The University of Arizona  
Engineering Building, 111  
PO.Box 210020  
Tucson, AZ 85721-0020, USA  
szidar@sie.arizona.edu  
Numerical Methods, Game Th., Dynamic Systems,

Institute of Mathematics  
 Gerhard Mercator University  
 D-47048 Duisburg  
 Germany  
 tel.0049-203-379-2676  
 knoop@math.uni-duisburg.de  
 Approximation Theory, Interpolation

16) Jerry Koliha  
 Dept. of Mathematics & Statistics  
 University of Melbourne  
 VIC 3010, Melbourne  
 Australia  
 koliha@unimelb.edu.au  
 Inequalities, Operator Theory,  
 Matrix Analysis, Generalized Inverses

17) Mustafa Kulenovic  
 Department of Mathematics  
 University of Rhode Island  
 Kingston, RI 02881, USA  
 kulenm@math.uri.edu  
 Differential and Difference Equations

18) Gerassimos Ladas  
 Department of Mathematics  
 University of Rhode Island  
 Kingston, RI 02881, USA  
 gladas@math.uri.edu  
 Differential and Difference Equations

19) V. Lakshmikantham  
 Department of Mathematical Sciences  
 Florida Institute of Technology  
 Melbourne, FL 32901  
 e-mail: lakshmik@fit.edu  
 Ordinary and Partial Differential  
 Equations,  
 Hybrid Systems, Nonlinear Analysis

20) Rupert Lasser  
 Institut für Biomathematik & Biomertie, GSF  
 -National Research Center for environment  
 and health  
 Ingolstaedter landstr.1  
 D-85764 Neuherberg, Germany  
 lasser@gsf.de  
 Orthogonal Polynomials, Fourier Analysis,  
 Mathematical Biology

Multicriteria Decision making,  
 Conflict Resolution, Applications  
 in Economics and Natural Resources  
 Management

35) Gancho Tachev  
 Dept. of Mathematics  
 Univ. of Architecture, Civil Eng. and Geodesy  
 1 Hr. Smirnenski blvd  
 BG-1421 Sofia, Bulgaria  
 gtt\_fte@uacg.bg  
 Approximation Theory

36) Manfred Tasche  
 Department of Mathematics  
 University of Rostock  
 D-18051 Rostock  
 Germany  
 manfred.tasche@mathematik.uni-rostock.de  
 Approximation Th., Wavelet, Fourier Analysis,  
 Numerical Methods, Signal Processing,  
 Image Processing, Harmonic Analysis

37) Chris P. Tsokos  
 Department of Mathematics  
 University of South Florida  
 4202 E. Fowler Ave., PHY 114  
 Tampa, FL 33620-5700, USA  
 profcpt@math.usf.edu, profcpt@chumal.cas.usf.edu  
 Stochastic Systems, Biomathematics,  
 Environmental Systems, Reliability Th.

38) Lutz Volkmann  
 Lehrstuhl II für Mathematik  
 RWTH-Aachen  
 Templergraben 55  
 D-52062 Aachen  
 Germany  
 volkm@math2.rwth-aachen.de  
 Complex Analysis, Combinatorics, Graph Theory

# On the numerical solution of nonlinear delay differential equations

S. Karimi Vanani and A. Aminataei

*Department of Mathematics, Faculty of Science,*

*K. N. Toosi University of Technology,*

*P.O. Box: 16315 – 1618, Tehran, Iran.*

*e-mail addresses: ataiei@kntu.ac.ir*

and

*solatkarimi@yahoo.com*

## Abstract

In this paper, we present a method to solve the delay differential equations. First, we convert the delay differential equations to power series, then we transform the power series into Padé series form, which gives an arbitrary order for solving delay differential equations. The advantages of using of the proposed method are presented. In the sequel, presented numerical solutions of some experiments and comparing with several other methods illustrate the high accuracy and efficiency of the proposed method. Also, the experiments show that the method works well for non-linear and problems on large intervals, too.

**Keywords:** Delay differential equations; Arbitrary order; Power series; Padé series.

**2000 Mathematics Subject Classification:** 65N; 65L10; 65N55.

## 1 Introduction

Delay differential equations (DDEs), arise in many areas of various mathematical modeling. For instance, infectious diseases, population dynamics, physiological and pharmaceutical kinetics and chemical kinetics, the navigational control of ships and aircraft and control problems. There are many books on the application of DDEs which we can point out to the books of Driver [5], Gopalsamy [8], Halanay [9] and Kuang [10]. Many different methods have been presented for numerical solution of DDEs such that we can point out to the Bellman's method of steps [3], waveform relaxation method [6], Runge-Kutta method and continuous Runge-Kutta method ([1] and [2]), spline methods ([11] and [12]) and Adomian decomposition method [13].

Out of these methods, we desire to solve the following DDE by the proposed method.

$$\begin{aligned} y'(t) &= f(t, y(t), y(t - \tau_1(t, y(t))), \dots, y(t - \tau_m(t, y(t)))), \quad t \geq t_0, \\ y(t) &= \phi(t), \quad t \leq t_0, \end{aligned} \tag{1}$$

where  $f : [t_0, \infty) \times \mathbb{R}^{m+1} \rightarrow \mathbb{R}$  is a smooth function and  $\tau_i(t, y(t)), i = 1, 2, \dots, m$ , are continuous delay functions on  $[t_0, \infty) \times \mathbb{R}$ . Also  $\phi(t)$  represents the initial function or the initial data.

The proposed method is easy with high accuracy and efficiency. Also, this method produces an approximate polynomial on the interval only in a few terms. For instance, in order to discuss about the efficiency of the proposed method we state its advantage with respect to Runge-Kutta method which is the most important method for numerical solution of DDEs. Runge-Kutta method like as all discrete variable methods produce approximations  $y_n$  to  $y(t_n)$  on a mesh  $t_n$  in the interval and starting with the given initial value,  $y_0 = y(t_0)$  at  $t_0$  and stepping from  $y_n \approx y(t_n)$  a distance of  $h_n$  to  $y_{n+1} \approx y(t_{n+1})$  at  $t_{n+1} = t_n + h_n$ . The step size  $h_n$  is chosen as small as necessary to get an accurate approximation. It is chosen as big as possible so as to reach the end of the interval in as steps as possible. In the original form, Runge-Kutta method produces answers only at mesh points, but our method produces an approximate polynomial on the desired intervals which gives the desired accuracy in a few terms, only. Thus the computations will be decreased and the main difference between this method and Runge-Kutta method is observable.

The organization of this paper is as follows: Section 2 is devoted to introduce the proposed method and then we transform the approximate polynomial gained by the proposed method to Padé approximate series. In section 3, we consider the error estimation of the method and In section 4, we present some experiments in-which their numerical results illustrate the high accuracy and efficiency of the method for non-linear and problems on large intervals. Finally, the last Section, consider some conclusions.

## 2 The proposed method

For solving equation (1), without loss of generality we assume that  $t_0 = 0$  (every interval  $[t_0, \infty)$  can be mapped to the  $[0, \infty)$ , easily).

First suppose that  $\{t - \tau_j(t, y(t))\}_{j=0}^m$  can be expanded as

$$t - \tau_j(t, y(t)) = \sum_{k=0}^{s_j} \beta_{kj}(t) y^k(t), \quad j = 0, 1, \dots, m, \quad (2)$$

where  $\tau_{kj}(t)$  are polynomial functions. Also, suppose that  $f(t, y(t), y(t - \tau_1(t, y(t))), \dots, y(t - \tau_m(t, y(t))))$  can be expanded as

$$f(t, y(t), y(t - \tau_1(t, y(t))), \dots, y(t - \tau_m(t, y(t)))) = \sum_{i_0=0}^{n_0} \sum_{i_1=0}^{n_1} \dots \sum_{i_m=0}^{n_m} \beta_{i_0, i_1, \dots, i_m}(t) y^{i_0}(t) y^{i_1}(t - \tau_1(t, y(t))) \dots y^{i_m}(t - \tau_m(t, y(t))), \quad (3)$$

where  $\beta_{i_0, i_1, \dots, i_m}(t)$  are polynomial functions. Now, we assumed that  $y_0 = y(0) = \phi(0)$  and the approximate solution is

$$\tilde{y}_1(t) = y_0 + et, \quad (4)$$

where  $e$  is a coefficient which obtain as follows:

We substitute (4) in (1) and by implementing (2) and (3) we gain a polynomial to approximate the expression

$$f(t, y_0 + et, y_0 + e(t - \tau_1(t, y_0 + et)), \dots, y_0 + e(t - \tau_m(t, y_0 + et))). \quad (5)$$

Suppose the approximate polynomial is in the form of

$$f_0 + f_1 t + f_2 t^2 + \dots + f_k t^k + \dots + f_M t^M, \quad (6)$$

where  $M$  is the maximum power of  $t$  in computations. Also,  $\tilde{y}'(t) = e$ . Thus, we obtain an approximation of equation (1) as

$$e = f_0 + f_1 t + f_2 t^2 + \dots + f_k t^k + \dots + f_M t^M, \quad (7)$$

neglecting of higher order terms, we have

$$e - f_0 + O(t) = 0, \quad (8)$$

therefore, we obtain  $e = f_0$ . Let us suppose that  $e = y_1$ . Thus, we have the following approximation of order one,

$$\tilde{y}_1(t) = y_0 + y_1 t. \quad (9)$$

In next step, we assume that new approximate solution is

$$\tilde{y}_2(t) = y_0 + y_1 t + et^2. \quad (10)$$

In the same way and neglecting of higher order terms (here  $O(t^2)$ ), the value of  $e$  will be obtained. Repeating the above procedure for aforesaid term and higher terms, we can get the arbitrary order power series of the solutions for equation (1) as

$$\tilde{y}_n(t) = y_0 + y_1 t + y_2 t^2 + \dots + y_n t^n. \quad (11)$$

Also, the power series given by the above procedure can be transformed into Padé series, easily.

Generally, Suppose that we are given a power series  $\sum_{i=0}^{\infty} a_i t^i$ , representing a function  $f(t)$ , so that

$$f(t) = \sum_{i=0}^{\infty} a_i t^i. \quad (12)$$

A Padé approximate is a rational fraction

$$[L/M] = \frac{p_0 + p_1 t + \dots + p_L t^L}{q_0 + q_1 t + \dots + q_M t^M}, \quad (13)$$

which has a Maclaurin expansion which agrees with (12) as far as possible. Notice that in (13) there are  $L + 1$  numerator coefficients and  $M + 1$  denominator coefficients. There is a

more or less irrelevant common factor between them, and for definiteness we take  $q_0 = 1$ . This choice turns out to be an essential part of the precise definition and (13) is our conventional notation with this choice for  $q_0$ . So there are  $L + 1$  independent numerator coefficients and  $M$  independent denominator coefficients, making  $L + M + 1$  unknown coefficients in all. This number suggests that normally the  $[L/M]$  ought to fit the power series (12) through the orders  $1, t, t^2, \dots, t^{L+M}$  in the notation of formal power series,

$$\sum_{i=0}^n a_i t^i = \frac{p_0 + p_1 t + \dots + p_L t^L}{q_0 + q_1 t + \dots + q_M t^M} + O(t^{L+M+1}). \quad (14)$$

Multiply the both side of (14) by the denominator of right side in (14) and compare the coefficients of both sides in (14). We have

$$a_l + \sum_{k=1}^M a_{l-k} q_k = p_l, \quad l = 0, 1, \dots, M, \quad (15)$$

$$a_l + \sum_{k=1}^L a_{l-k} q_k = 0, \quad l = M + 1, \dots, M + L. \quad (16)$$

Solving the linear equation in (16), we have  $q_k, k = 1, \dots, L$ ; and substituting  $q_k$  into (15), we obtain  $p_l$  for all  $l = 0, \dots, M$ , such as [4]. Therefore, we have constructed a  $[L/M]$  Padé approximation, which agrees with  $\sum_{i=0}^{\infty} a_i t^i$  through order  $t^{L+M}$ . If  $M \leq L \leq M+2$ , where  $M$  and  $L$  are the degree of numerator and denominator in Padé series, respectively, then Padé series gives an A-stable formula for an ordinary differential equation [14].

### 3 Error estimation

In this section, an error estimation for the approximate solution of (1) is obtained. Let us call  $e_n(t) = |y(t) - \tilde{y}_n(t)|$  as the error function of the approximate solution  $\tilde{y}(t)$  to  $y(t)$ , where  $y(t)$  is the exact solution of (1). Hence,  $\tilde{y}(t)$  satisfies the following problem:

$$\tilde{y}'_n(t) = f(t, \tilde{y}_n(t), \tilde{y}_n(t - \tau_1(t, \tilde{y}_n(t))), \dots, \tilde{y}_n(t - \tau_m(t, \tilde{y}_n(t)))) + K_n(t), \quad t \geq t_0. \quad (17)$$

The perturbation term  $K_n(t)$  can be obtained by substituting the computed solution  $\tilde{y}_n(t)$  into the equation

$$K_n(t) = \tilde{y}'_n(t) - f(t, \tilde{y}_n(t), \tilde{y}_n(t - \tau_1(t, \tilde{y}_n(t))), \dots, \tilde{y}_n(t - \tau_m(t, \tilde{y}_n(t))))), \quad t \geq t_0. \quad (18)$$

We proceed to find an approximation  $\tilde{e}_n(t)$  to the error function  $e_n(t)$  in the same way as we did before for the solution of problem (2). Subtracting (18) from (1), the error function  $e_n(t)$  satisfies the problem.

$$e_n(t) - f(t, e_n(t), e(t - \tau_1(t, e_n(t))), \dots, e(t - \tau_m(t, e_n(t)))) = -K_n(t), \quad t \geq t_0. \quad (19)$$

It should be noted that in order to construct the approximate  $\tilde{e}_n(t)$  to  $e_n(t)$ , only the equation (19) needs to be recomputed in the same way as we did before for the solution of problem (1).

## 4 Numerical experiments

In this section, we present four experiments in-which their numerical results state the high accuracy and efficiency of the proposed method. Problems 4.1 and 4.2 are chosen so simple, but Problems 4.3 and 4.4 are nonlinear and almost complicated. In these problems we observe the high accuracy and efficiency of the method, obviously.

**Problem 4.1** Consider the following DDE [11, 12, 13],

$$\begin{cases} y'(t) = \frac{1}{2}e^{\frac{t}{2}}y(\frac{t}{2}) + \frac{1}{2}y(t), & 0 \leq t \leq 1, \\ y(0) = 1. \end{cases}$$

The exact solution is  $y(t) = e^t$ .

We have obtained 17 terms of the approximate solution. Comparing the results with other methods [11, 12, 13], we found that Padé approximate solution is the best as shown in Table 1.

**Table 1: Comparison of errors between some methods and Padé approximate solution**

$t_i$	$h = 0.001$ [12]	$h = 0.001$ [11]	ADM [13]	Padé approximate solution
0.1	$8.23 \times 10^{-12}$	$1.12 \times 10^{-15}$	$1.30 \times 10^{-15}$	0.00
0.2	$1.37 \times 10^{-11}$	$3.10 \times 10^{-15}$	$1.40 \times 10^{-15}$	$1.00 \times 10^{-15}$
0.3	$2.31 \times 10^{-11}$	$4.81 \times 10^{-15}$	$1.50 \times 10^{-15}$	$1.00 \times 10^{-15}$
0.4	$3.27 \times 10^{-11}$	$7.54 \times 10^{-15}$	$1.90 \times 10^{-15}$	$1.00 \times 10^{-15}$
0.5	$4.48 \times 10^{-11}$	$9.73 \times 10^{-15}$	$2.10 \times 10^{-15}$	0.00
0.6	$5.86 \times 10^{-11}$	$1.39 \times 10^{-14}$	$2.20 \times 10^{-15}$	0.00
0.7	$7.43 \times 10^{-11}$	$1.76 \times 10^{-14}$	$2.40 \times 10^{-15}$	$1.00 \times 10^{-15}$
0.8	$9.54 \times 10^{-11}$	$2.13 \times 10^{-14}$	$2.60 \times 10^{-15}$	0.00
0.9	$9.87 \times 10^{-11}$	$2.84 \times 10^{-14}$	$2.20 \times 10^{-15}$	$2.00 \times 10^{-15}$
1.0	$1.43 \times 10^{-10}$	$3.19 \times 10^{-14}$	$2.60 \times 10^{-15}$	$2.00 \times 10^{-15}$

Also, during the running of program we find out the run time of Padé approximation is 0.451 second and run time of the Adomian decomposition method is 12.047 seconds. Thus in the case of numerical solution of DDEs, we prefer Padé approximation to the Adomian decomposition method.

**Problem 4.2** Consider the following DDE [13],

$$\begin{cases} y'(t) = 1 - 2y^2(\frac{t}{2}), & 0 \leq t \leq 1, \\ y(0) = 0. \end{cases}$$

The exact solution is  $y(t) = \sin(t)$ .

The approximate solution is obtained with  $n = 16$  is given in Table 2.



**Table 2:** Exact solution, Padé approximate solution and the error value of  $y(t)$ 

$t_i$	$y(t_i)$	$\tilde{y}(t_i)$	$ y(t_i) - \tilde{y}(t_i) $
0.0	1	1.0000000000000000	0
0.1	0.0998334166468281	0.0998334166468281	0
0.2	0.1986693307950612	0.1986693307950612	0
0.3	0.2955202066613396	0.2955202066613396	0
0.4	0.3894183423086505	0.3894183423086505	0
0.5	0.4794255386042030	0.4794255386042030	0
0.6	0.5646424733950354	0.5646424733950354	0
0.7	0.6442176872376911	0.6442176872376910	$1 \times 10^{-16}$
0.8	0.7173560908995228	0.7173560908995227	$1 \times 10^{-16}$
0.9	0.7833269096274834	0.7833269096274829	$5 \times 10^{-16}$
1.0	0.8414709848078965	0.8414709848078937	$2.2 \times 10^{-15}$

In the two previous experiments you observed that, this method of solution produces an approximate solution in minimum number of terms and acts accurate. This exhibits one of the main advantages of the method in spite of its simplicity.

**Problem 4.3** Consider the following DDE,

$$\begin{cases} y'(t) + \sin(t)y^2(t) + e^t y(\frac{y(t)}{2})y(t - \sin(t)) = 2t\sin(t) + t^2\cos(t) + \\ \quad + \sin(t)(t^2\sin(t))^2 + e^t(\frac{t^2\sin(t)}{2})^2\sin(\frac{t^2\sin(t)}{2})(t - \sin(t))^2\sin(t - \sin(t)), & t \geq 0, \\ y(t) = t^2\sin(t), & t \leq 0. \end{cases}$$

The considered delays are

$$\tau_1(t, y(t)) = t - \frac{y(t)}{2}, \tau_2(t, y(t)) = \sin(t).$$

The exact solution is  $y(t) = t^2\sin(t)$ . The approximate solution is obtained with  $n = 52$ , and the results in the interval  $[0, 10]$  are given in Table 3.

Table 3: Exact solution, Padé approximate solution and the error value of  $y(t)$ 

$t_i$	$y(t_i)$	$\tilde{y}(t_i)$	$ y(t_i) - \tilde{y}(t_i) $
0	0	0.000000000000000	0
.5	0.119856384651050	0.119856384651050	0
1	0.841470984807896	0.841470984807896	0
1.5	2.244363719859122	2.244363719859122	0
2	3.637189707302727	3.637189707302727	0
2.5	3.740450900649728	3.740450900649728	0
3	1.270080072538805	1.270080072538805	0
3.5	-4.297094539197843	-4.297094539197843	0
4	-12.10883992492685	-12.10883992492685	0
4.5	-19.79498488271822	-19.79498488271822	0
5	-23.97310686657846	-23.97310686657846	0
5.5	-21.34259484850435	-21.34259484850436	$1 \times 10^{-15}$
6	-10.05895793516133	-10.05895793516133	0
6.5	9.088819496710205	9.088819496710206	$1 \times 10^{-15}$
7	32.19234333722067	32.19234333722067	0
7.5	52.76249869357906	52.76249869357906	0
8	63.31892778389644	63.31892778389643	$1 \times 10^{-15}$
8.5	57.69069388704717	57.69069388704717	0
9	33.38159730458228	33.38159730458228	0
9.5	-6.782388621678290	-6.782388621678286	$4 \times 10^{-15}$
10	-54.40211108893698	-54.40211108893692	$6 \times 10^{-15}$

In this experiment, the efficiency of the method on large intervals is observable.

**Problem 4.4** Consider the following DDE,

$$\begin{cases} y'(t) + y^3(y(t)) + y(y^3(t)) = 2te^{-2t} - 2t^2e^{-2t} + ((t^2e^{-2t})^2e^{-2(t^2e^{-2t})})^3 + \\ \quad + ((t^2e^{-2t})^3)^2e^{-2(t^2e^{-2t})^3}, \quad t \geq 0, \\ y(t) = t^2e^{-2t}, \quad t \leq 0. \end{cases}$$

The considered delays are

$$\tau_1(t, y(t)) = y(t), \tau_2(t, y(t)) = y^3(t).$$

The exact solution is  $y(t) = t^2e^{-2t}$ .

The approximate solution is obtained with  $n = 40$ , and the results in the interval  $[0, e]$  are given in Table 4.

**Table 4:** Exact solution, Padé approximate solution and the error value of  $y(t)$ 

$t_i$	$y(t_i)$	$\tilde{y}(t_i)$	$ y(t_i) - \tilde{y}(t_i) $
0	0	0.000000000000000	0
.1e	0.04290244113814	0.04290244113814	0
.2e	0.09964030214245	0.09964030214245	0
.3e	0.13016990689938	0.13016990689938	0
.4e	0.13436343801038	0.13436343801038	0
.5e	0.12189732467981	0.12189732467982	$1 \times 10^{-14}$
.6e	0.10191772165232	0.10191772165233	$1 \times 10^{-14}$
.7e	0.08054458091031	0.08054458091034	$3 \times 10^{-14}$
.8e	0.06108200165303	0.06108200165308	$5 \times 10^{-14}$
.9e	0.04488604552887	0.04488604552886	$1 \times 10^{-14}$
e	0.03217506012167	0.03217506012167	0

## 5 Conclusions

We present a method for solving delay differential equations. This method is easy to implement and yields desired accuracy in minimum number of terms, such that numerical results demonstrate this issue. Comparing this method with several other methods shows that the method is more efficient and accurate. We also observed that the method works excellently for nonlinear delay differential equations. Thus the proposed method is suggested as an efficient method for the numerical solution of delay differential equations. The computations associated with the experiments discussed above, were performed in Maple 10 on PC, CPU 2.4 GHz.

## References

- [1] A. Bellen and M. Zennaro, Adaptive integration of delay differential equations, In *Proceeding of The Workshop CNRS-NSF: Advances in Time Delay Systems*. Paris, 2003.
- [2] A. Bellen and M. Zennaro, Numerical methods for delay differential equations, numerical mathematics and scientific computations series, *Oxford University Press, Oxford*, 2003.
- [3] R. Bellman, On the computational solution of differential difference equations, *J. Math. Anal. Appl.* 2., 108-110, 1961.
- [4] E. Çelik, E. Karaduman, and M. Bayram, A numerical method to solve chemical differential algebraic equations, *Int. J. Quantum Chem.* 89., 447-451, 2002.
- [5] R.D. Driver, Ordinary and Delay Differential Equations, *Applied Mathematics Series* 20, 1977.

- [6] A. Ruehli, E. Lelarsmee and A. Sangiovanni-Vincentelli, The waveform relaxation method for time domain analysis in large scale integrated circuits, *IEEE Trans. CAD IC Syst* 1., 131-145, 1982.
- [7] E.L. El'sgol'ts and S.B. Norkin, Introduction to the Theory and Applications of Differential Equations with Deviating Arguments, *Mathematics in Science and Engineering* 105., 1973.
- [8] K. Gopalsamy, Stability and oscillations in delay differential equations of population dynamics, *Technical Report, Kluwer, Dordrecht, The Netherlands*, 1992.
- [9] A. Halanay, Differential Equations: Stability, Oscillation, Time lags, *Mathematics in Science and Engineering* 23, 1966.
- [10] Y. Kuang, Delay Differential Equations with Applications in Population Dynamics, *Mathematics in Science and Engineering* 191., 1993.
- [11] A. El-Safty, M.S. Salim and M.A. El-Khatib, Convergence of the spline function for delay dynamic system, *International Journal of Computer Mathematics*, 80(4)., 509-518, 2003.
- [12] M. Shadia, Numerical solution of delay differential and neutral differential equations using spline methods, *PhD thesis, Assuit University*, 1992.
- [13] D.J. Evans and K.R. Raslan, The Adomian decomposition method for solving delay differential equation, *International Journal of Computer Mathematics*, 82(1)., 49-54, 2005.
- [14] E. Çelik, E. Karaduman, and M. Bayram, Numerical solutions of chemical differential algebraic equations, *Applied Mathematics and Computation* 139., 259-264, 2003.

# Inclusion Theorems for Absolute Matrix Summability Methods

W. T. Sulaiman

Department of Computer Engineering, College of Engineering,  
University of Mosul, Iraq.

## Abstract

In this paper, a general theorem concerning  $\varphi - |T|_k$  factors of infinite series has been proved

## 1. Introduction

Let  $(\varphi_n)$  be a sequence of positive real numbers, let  $\sum a_n$  be an infinite series with the sequence of partial sums  $(s_n)$ . Let  $(u_n)$  denote the  $n$ -th  $(C,1)$  means of the sequence  $(na_n)$ . The series  $\sum a_n$  is said to be summable  $|C,1|_k$ ,  $k \geq 1$ , if (see[1])

$$(1.1) \quad \sum_{n=1}^{\infty} \frac{1}{n} |u_n|^k < \infty.$$

and it is said to be summable  $\varphi - |C,1|_k$ ,  $k \geq 1$ , if (see [5])

$$(1.2) \quad \sum_{n=1}^{\infty} \frac{\varphi_n^{k-1}}{n^k} |u_n|^k < \infty.$$

If we are taking  $\varphi_n = n$ ,  $\varphi - |C,1|_k$  reduces to  $|C,1|_k$  summability.

Let  $(p_n)$  be a sequence of positive numbers such that

$$P_n = \sum_{v=0}^n p_v \rightarrow \infty \text{ as } n \rightarrow \infty \quad (p_{-1} = P_{-1} = 0).$$

The sequence-to-sequence transformation

$$(1.3) \quad v_n = \frac{1}{P_n} \sum_{v=0}^n p_v s_v$$

defines the sequence  $(v_n)$  of the Riesz mean or simply the  $(\bar{N}, p_n)$  mean of the sequence  $(s_n)$  generated by the sequence of coefficients  $(p_n)$  (see[2]). The series

$\sum a_n$  is said to be summable  $|R, p_n|_k$ ,  $k \geq 1$  if

$$(1.4) \quad \sum_{n=1}^{\infty} n^{k-1} |v_n - v_{n-1}|^k < \infty.$$

In the special case when  $p_n = 1$  for all  $n$ , then  $|R, p_n|_k$  summability is the same as  $|C,1|_k$  summability. The series  $\sum a_n$  is summable  $\varphi - |R, p_n|_k$ ,  $k \geq 1$ , if

$$\sum_{n=1}^{\infty} \varphi_n^{k-1} |v_n - v_{n-1}|^k < \infty.$$

For  $\varphi_n = n$ ,  $\varphi - |R, p_n|_k$  summability is the same as  $|R, p_n|_k$  summability .

For arbitrary lower triangular matrix  $T = (t_{nv})$ , the series  $\sum a_n$  is summable  $|T|_k$ ,  $k \geq 1$ , if

$$(1.5) \quad \sum_{n=1}^{\infty} n^{k-1} |\Delta t_{n-1}|^k < \infty ,$$

where

$$t_n = \sum_{k=0}^n t_{nk} s_k .$$

The series  $\sum a_n$  is summable  $\varphi - |T|_k$ ,  $k \geq 1$ , if

$$(1.6) \quad \sum_{n=1}^{\infty} \varphi_n^{k-1} |\Delta t_{n-1}|^k < \infty .$$

Two matrices  $\bar{T} = (\bar{t}_{nv})$  and  $\hat{T} = (\hat{t}_{nv})$  can be associated with  $T$  as follows :

The entries  $\bar{t}_{nv}$  and  $\hat{t}_{nv}$  are defined by

$$(1.7) \quad \bar{t}_{nk} = \sum_{v=k}^n t_{nv} , \quad \hat{t}_{nv} = \bar{t}_{n,v} - \bar{t}_{n-1,v} .$$

By above, we have

$$(1.8) \quad t_n = \sum_{k=0}^n t_{nk} s_k = \sum_{k=0}^n t_{nk} \sum_{v=0}^n a_v = \sum_{v=0}^n a_v \sum_{k=v}^n t_{nv} = \sum_{v=0}^n \bar{t}_{nv} a_v ,$$

$$(1.9) \quad Y_n := t_n - t_{n-1} = \sum_{v=0}^n \bar{t}_{nv} a_v - \sum_{v=0}^{n-1} \bar{t}_{n-1,v} a_v = \sum_{v=0}^n \hat{t}_{nv} a_v , \quad \text{as } \bar{t}_{n-1,n} = 0 .$$

Concerning  $|C, 1|_k$  summability, Mazhar [3] has proved the following

**Theorem 1.1.** *If*

$$(1.10) \quad \lambda_m = o(1), \quad \text{as } m \rightarrow \infty ,$$

$$(1.11) \quad \sum_{n=1}^m n \log n |\Delta^2 \lambda_n| = O(1), \quad \text{as } m \rightarrow \infty$$

$$(1.12) \quad \sum_{v=1}^m \frac{|t_v|^k}{v} = O(\log m) \quad \text{as } m \rightarrow \infty ,$$

then the series  $\sum a_n \lambda_n$  is summable  $|C, 1|_k$ ,  $k \geq 1$ .

Ozarslan [4], on the other hand , generalized the previous result by giving the following

**Theorem 1.2.** *Let  $(\varphi_n)$  be a sequence of positive real numbers and the conditions (1.10)-(1.11) of Theorem (1.10) are satisfied . If*

$$(1.13) \quad \sum_{v=1}^m \frac{\varphi_v^{k-1}}{v^k} |t_v|^k = O(\log m) \quad \text{as } m \rightarrow \infty ,$$

$$(1.14) \quad \sum_{n=v}^{\infty} \frac{\varphi_n^{k-1}}{n^{k+1}} = O\left(\frac{\varphi_v^{k-1}}{v^k}\right),$$

then the series  $\sum a_n \lambda_n$  is summable  $\varphi - |C, 1|_k$ ,  $k \geq 1$ .

It should be mentioned that on taking  $\varphi_n = n$  in Theorem (1.2), we get Theorem 1.1.

## 2. Main result

**Theorem 2.1.** Let  $(\varphi_n), (Z_n)$  be sequences of positive real numbers such that  $(Z_n)$  is non-decreasing and the condition (1.10), is satisfied. If

$$(2.1) \quad |t_{nn}| = O(p_n / P_n),$$

$$(2.2) \quad n \Delta Z_n = O(Z_n),$$

$$(2.3) \quad \sum_{n=1}^{\infty} n Z_n |\Delta^2 \lambda_n| < \infty,$$

$$(2.4) \quad \sum_{v=1}^n |t_{vv}| |\hat{t}_{n,v+1}| = O(|t_{nn}|),$$

$$(2.5) \quad \sum_{n=v}^m \varphi_n^{k-1} |t_{nn}|^{k-1} |\hat{t}_{nv}| = O\left((v |t_{vv}|)^{k-1}\right),$$

$$(2.6) \quad \sum_{v=1}^n |\Delta \hat{t}_{nv}| = O(|t_{nn}|),$$

$$(2.7) \quad \sum_{n=v}^m \varphi_n^{k-1} |t_{nn}|^{k-1} |\Delta \hat{t}_{nv}| = O\left(v^{k-1} |t_{vv}|^k\right),$$

$$(2.8) \quad \sum_{n=v}^m \varphi_n^{k-1} |\hat{t}_{n,v+1}|^k = O(1).$$

Then the series  $\sum a_n \lambda_n Z_n$  is summable  $\varphi - |T|_k$  whenever  $\sum a_n$  is summable  $|R, p_n|_k$ ,  $k \geq 1$ .

## 3. Lemma

The following Lemma is needed

**Lemma 3.1.** The conditions (1.10) and (2.4) implies

$$(3.1) \quad \sum_{n=1}^{\infty} Z_n |\Delta \lambda_n| = O(1),$$

$$(3.2) \quad \sum_{n=1}^{\infty} |\lambda_n| |\Delta Z_n| = O(1),$$

$$(3.3) \quad n Z_n |\Delta \lambda_n| = O(1), \quad \text{as } n \rightarrow \infty,$$

$$(3.4) \quad Z_n |\lambda_n| = O(1), \quad \text{as } n \rightarrow \infty.$$

**Proof.** By virtue of (1.10),

$$\begin{aligned} \sum_{n=1}^{\infty} Z_n |\Delta \lambda_n| &= \sum_{n=1}^{\infty} Z_n \sum_{v=n}^{\infty} \Delta |\Delta \lambda_v| \\ &= \sum_{v=1}^{\infty} |\Delta^2 \lambda_v| \sum_{n=1}^v Z_v \\ &= O(1) \sum_{v=1}^{\infty} v Z_v |\Delta^2 \lambda_v| \\ &= O(1). \end{aligned}$$

$$\begin{aligned} \sum_{n=1}^{\infty} |\lambda_n| |\Delta Z_n| &= \sum_{n=1}^{\infty} |\Delta Z_n| \sum_{v=n}^{\infty} |\Delta \lambda_v| \\ &\leq \sum_{n=1}^{\infty} |\Delta Z_n| \sum_{v=n}^{\infty} |\Delta \lambda_v| \\ &= O(1) \sum_{v=1}^{\infty} |\Delta \lambda_v| \sum_{n=1}^v |\Delta Z_n| \\ &= O(1) \sum_{v=1}^{\infty} |\Delta \lambda_v| \sum_{n=1}^v (Z_{n+1} - Z_n) \\ &= O(1) \sum_{v=1}^{\infty} |\Delta \lambda_v| (Z_{v+1} - Z_1) \\ &= O(1) \sum_{v=1}^{\infty} |\Delta \lambda_v| Z_v \\ &= O(1). \end{aligned}$$

$$\begin{aligned} n Z_n |\Delta \lambda_n| &= n Z_n \left| \sum_{v=n}^{\infty} \Delta |\Delta \lambda_v| \right| \\ &\leq n Z_n \sum_{v=n}^{\infty} |\Delta |\Delta \lambda_v|| \\ &\leq n Z_n \sum_{v=n}^{\infty} |\Delta^2 \lambda_v| \\ &= O(1) \sum_{n=v}^{\infty} v Z_v |\Delta^2 \lambda_v| \\ &= O(1). \end{aligned}$$

$$\begin{aligned} Z_n |\lambda_n| &= Z_n \sum_{v=n}^{\infty} \Delta |\lambda_v| \\ &\leq Z_n \sum_{v=n}^{\infty} |\Delta \lambda_v| \end{aligned}$$



$$\begin{aligned}
&= O(1) \sum_{v=n}^{\infty} Z_v |\Delta \lambda_v| \\
&= O(1), \text{ by the first part.}
\end{aligned}$$

#### 4. Proof of Theorem 2.1

Let  $(t_n)$  denote the  $(\bar{N}, p_n)$  mean of the series  $\sum a_n$ . By definition, we have

$$X_n := t_n - t_{n-1} = \frac{P_n}{P_n P_{n-1}} \sum_{v=1}^{n-1} P_{v-1} a_v.$$

Also, we write

$$T_n = \sum_{v=0}^n t_{nv} \sum_{i=0}^v a_i \lambda_i Z_i = \sum_{i=0}^n a_i \lambda_i Z_i \sum_{v=i}^n t_{nv} = \sum_{i=0}^n \bar{t}_{ni} a_i \lambda_i Z_i.$$

Therefore, we have

$$Y_n := T_n - T_{n-1} = \sum_{v=0}^n \hat{t}_{nv} a_v \lambda_v Z_v.$$

By Abel's transformation,

$$\begin{aligned}
Y_n &= \sum_{v=1}^n P_{v-1} a_v \frac{\hat{t}_{nv} \lambda_v Z_v}{P_{v-1}} \\
&= \sum_{v=1}^{n-1} \left( \sum_{r=1}^v P_{r-1} a_r \right) \Delta_v \left( \frac{\hat{t}_{nv} \lambda_v Z_v}{P_{v-1}} \right) + \left( \sum_{v=1}^n P_{v-1} a_v \right) \frac{t_{nn} \lambda_n Z_n}{P_{n-1}} \quad (\hat{t}_{nn} = t_{nn}) \\
&= \sum_{v=1}^{n-1} \frac{P_v P_{v-1}}{P_v} X_v \Delta_v \left( \frac{\hat{t}_{nv} \lambda_v Z_v}{P_{v-1}} \right) + \frac{P_n P_{n-1}}{P_n} X_n \frac{t_{nn} \lambda_n Z_n}{P_{n-1}} \\
&= \sum_{v=1}^{n-1} \left( X_v \hat{t}_{nv} \lambda_v Z_v + \frac{P_{v-1}}{P_v} X_v \Delta \hat{t}_{nv} \lambda_v Z_v + \frac{P_{v-1}}{P_v} X_v \hat{t}_{n,v+1} \Delta \lambda_v Z_v + \frac{P_{v-1}}{P_v} X_v \hat{t}_{n,v+1} \lambda_{v+1} \Delta Z_v \right) \\
&\quad + \frac{P_n}{P_n} X_n t_{nn} \lambda_n Z_n \\
&= Y_{n1} + Y_{n2} + Y_{n3} + Y_{n4} + Y_{n5}.
\end{aligned}$$

In order to prove the theorem, by Minkowski's inequality, it is sufficient to show that

$$\sum_{n=1}^{\infty} \phi_n^{k-1} |Y_{nr}|^k < \infty, \quad r = 1, 2, 3, 4, 5.$$

We now applying Holder's inequality

$$\begin{aligned}
\sum_{n=1}^m \phi_n^{k-1} |Y_{n1}|^k &= \sum_{n=1}^m \phi_n^{k-1} \left| \sum_{v=1}^{n-1} X_v \hat{t}_{nv} \lambda_v Z_v \right|^k \\
&= O(1) \sum_{n=1}^m \phi_n^{k-1} \sum_{v=1}^{n-1} |X_v|^k \frac{|\hat{t}_{nv}|^k}{|t_{vv}|^{k-1}} |\lambda_v|^k |Z_v|^k \left( \sum_{v=1}^{n-1} |t_{vv}| |\hat{t}_{nv}| \right)^{k-1}
\end{aligned}$$

$$\begin{aligned}
&= O(1) \sum_{n=1}^m \varphi_n^{k-1} |t_{nm}|^{k-1} \sum_{v=1}^{n-1} |X_v|^k \frac{|\hat{t}_{nv}|}{|t_{vv}|^{k-1}} |\lambda_v|^k |Z_v|^k \\
&= O(1) \sum_{v=1}^m |X_v|^k \frac{1}{|t_{vv}|^{k-1}} |\lambda_v|^k |Z_v|^k \sum_{n=v}^m \varphi_n^{k-1} |t_{nm}|^{k-1} |\hat{t}_{nv}| \\
&= O(1) \sum_{v=1}^m v^{k-1} |X_v|^k \\
&= O(1) .
\end{aligned}$$

$$\begin{aligned}
\sum_{n=1}^m \varphi_n^{k-1} |Y_{n2}|^k &= \sum_{n=1}^m \varphi_n^{k-1} \left| \sum_{v=1}^{n-1} \frac{P_{v-1}}{P_v} X_v \Delta_v \hat{t}_{nv} \lambda_v Z_v \right|^k \\
&= O(1) \sum_{n=1}^m \varphi_n^{k-1} \sum_{v=1}^{n-1} \frac{P_{v-1}^k}{P_v^k} |X_v|^k |\hat{\Delta}_v t_{nv}| |\lambda_v|^k Z_v^k \left( \sum_{v=1}^{n-1} |\Delta_v \hat{t}_{nv}| \right)^{k-1} \\
&= O(1) \sum_{n=1}^m \varphi_n^{k-1} |t_{nm}|^{k-1} \sum_{v=1}^{n-1} \frac{P_{v-1}^k}{P_v^k} |X_v|^k |\hat{\Delta}_v t_{nv}| |\lambda_v|^k Z_v^k \\
&= O(1) \sum_{v=1}^m \frac{P_v^k}{P_v^k} |X_v|^k |\lambda_v|^k Z_v^k \sum_{n=v}^m \varphi_n^{k-1} |t_{nm}|^{k-1} |\Delta_v \hat{t}_{nv}| \\
&= O(1) \sum_{v=1}^m \frac{P_v^k}{P_v^k} |X_v|^k v^{k-1} |t_{vv}|^{k-1} \\
&= O(1) \sum_{v=1}^m v^{k-1} |X_v|^k \\
&= O(1) .
\end{aligned}$$

$$\begin{aligned}
\sum_{n=1}^m \varphi_n^{k-1} |Y_{n3}|^k &= \sum_{n=1}^m \varphi_n^{k-1} \left| \sum_{v=1}^{n-1} \frac{P_{v-1}}{P_v} X_v \hat{t}_{n,v+1} \Delta \lambda_v Z_v \right|^k \\
&= O(1) \sum_{n=1}^m \varphi_n^{k-1} \sum_{v=1}^{n-1} \frac{P_{v-1}^k}{P_v^k} |X_v|^k |\hat{t}_{n,v+1}|^k |\Delta \lambda_v| Z_v \left( \sum_{v=1}^{n-1} |\Delta \lambda_v| Z_v \right)^{k-1} \\
&= O(1) \sum_{v=1}^m \frac{P_v^k}{P_v^k} |X_v|^k |\Delta \lambda_v| Z_v \sum_{n=v}^m \varphi_n^{k-1} |\hat{t}_{n,v+1}|^k \\
&= O(1) \sum_{v=1}^m v^{k-1} |X_v|^k \cdot v |\Delta \lambda_v| Z_v \\
&= O(1) \sum_{v=1}^{m-1} \left( \sum_{r=1}^v r^{k-1} |X_r|^k \right) \Delta(v |\Delta \lambda_v| Z_v) + O(1) \left( \sum_{v=1}^m v^{k-1} |X_v|^k \right) m |\Delta \lambda_m| Z_m \\
&= O(1) \sum_{v=1}^{m-1} |\Delta \lambda_v| Z_v + O(1) \sum_{v=1}^{m-1} (v+1) |\Delta Z_v| |\Delta \lambda_v| \\
&\quad + O(1) \sum_{v=1}^{m-1} (v+1) Z_{v+1} |\Delta^2 \lambda_v| + O(1)
\end{aligned}$$

$$\begin{aligned}
&= O(1) + O(1) \sum_{v=1}^{m-1} Z_v |\Delta \lambda_v| + O(1) \sum_{v=1}^{m-1} v Z_v |\Delta^2 \lambda_v| + O(1) \\
&= O(1) .
\end{aligned}$$

$$\begin{aligned}
\sum_{n=1}^m \varphi_n^{k-1} |Y_{n4}|^k &= \sum_{n=1}^m \varphi_n^{k-1} \left| \sum_{v=1}^{n-1} \frac{P_{v-1}}{P_v} X_v \hat{t}_{n,v+1} \lambda_{v+1} \Delta Z_v \right|^k \\
&= O(1) \sum_{n=1}^m \varphi_n^{k-1} \left| \sum_{v=1}^{n-1} \frac{P_{v-1}}{P_v} X_v \hat{t}_{n,v+1} \lambda_v \Delta Z_v \right|^k \\
&= O(1) \sum_{n=1}^m \varphi_n^{k-1} \sum_{v=1}^{n-1} \frac{P_v^k}{P_v^k} |X_v|^k |\hat{t}_{n,v+1}|^k |\lambda_v| |\Delta Z_v| \left( \sum_{v=1}^{n-1} |\lambda_v| |\Delta Z_v| \right)^{k-1} \\
&= O(1) \sum_{v=1}^m \frac{P_v^k}{P_v^k} |X_v|^k |\lambda_v| |\Delta Z_v| \sum_{v=1}^{n-1} \varphi_n^{k-1} |\hat{t}_{n,v+1}|^k \\
&= O(1) \sum_{v=1}^m v^{k-1} |X_v|^k |\lambda_v| Z_v \\
&= O(1) \sum_{v=1}^m v^{k-1} |X_v|^k \\
&= O(1) .
\end{aligned}$$

$$\begin{aligned}
\sum_{v=1}^m \varphi_n^{k-1} |Y_{n5}|^k &= \sum_{n=1}^m \varphi_n^{k-1} \left| \frac{P_n}{P_n} X_n t_{nn} \lambda_n Z_n \right|^k \\
&= O(1) \sum_{n=1}^m \varphi_n^{k-1} \frac{P_n^k}{P_n^k} |X_n|^k |t_{nn}|^k |\lambda_n|^k Z_n^k \\
&= O(1) \sum_{n=1}^m n^{k-1} |X_n|^k \\
&= O(1) .
\end{aligned}$$

## 5. Applications

**Corollary 5.1.** Let  $(\varphi_n), (Z_n)$  be sequences of positive real numbers such that  $(Z_n)$  is non-decreasing and let the conditions (1.10), (2.2) and (2.3) are satisfied. If

$$(5.1) \quad \sum_{n=v}^{\infty} \frac{\varphi_n^{k-1} P_n^k}{P_n^k P_{n-1}} = O \left( \frac{v^{k-1} \left( \frac{P_v}{P_v} \right)^{k-1}}{P_{v-1} \left( \frac{P_v}{P_v} \right)} \right),$$

$$(5.2) \quad \sum_{n=v}^{\infty} \frac{\varphi_n^{k-1} P_n^k}{P_n^k P_{n-1}} = O \left( \frac{v^{k-1} \left( \frac{P_v}{P_v} \right)^k}{p_v \left( \frac{P_v}{P_v} \right)} \right),$$

$$(5.3) \quad \sum_{n=v}^m \varphi_n^{k-1} \left( \frac{P_n}{P_n P_{n-1}} \right)^k = O \left( 1 / P_v^k \right).$$

Then the series  $\sum a_n \lambda_n Z_n$  is summable  $\varphi - |R, p_n|_k$  whenever  $\sum a_n$  is summable  $|R, p_n|_k$ ,  $k \geq 1$ .

**Proof.** Follows from theorem 2.1 by putting

$$t_{nv} = \frac{p_v}{P_n}, \quad \hat{t}_{nv} = \frac{p_n P_{v-1}}{P_n P_{n-1}}, \quad \Delta \hat{t}_{nv} = \frac{p_n p_v}{P_n P_{n-1}}.$$

**Corollary 5.2.** Let  $(\varphi_n), (Z_n)$  be sequences of positive real numbers such that  $(Z_n)$  is non-decreasing and the conditions (1.10), (2.2) and (2.3) are satisfied. If

$$(5.4) \quad \sum_{n=v}^{\infty} \frac{\varphi_n^{k-1}}{n^{k+1}} = O(1/v),$$

then the series  $\sum a_n \lambda_n Z_n$  is summable  $\varphi - |C, 1|_k$ , whenever  $\sum a_n$  is summable  $|C, 1|_k$ ,  $k \geq 1$ .

**Proof.** Follows from Corollary 5.1, by putting  $p_n = 1$  for all  $n$ .

## References

- [1] T. M. Flett, On an extension of absolute summability and some theorems of Littlewood and Paley, Proc. London. Math. Soc. 7 (1957), 113-141.
- [2] G. H. Hardy, Divergent series, Oxford Univ. Press. Oxford. 1949.
- [3] S. M. Mazhar, On  $|C, 1|_k$  summability factors of infinite series, Indian J. Math. 14 (1972), 45-48.
- [4] H. S. Ozarslan, On absolute Cesaro summability factors of infinite series, Communications in Mathematical Analysis 3 (2007), 53-56.
- [5] H. Seyhan, The absolute summability methods. Ph.D. Thesis, Kayseri (1995), 1-57.

# Integral Inequalities Concerning Triple Integrals

W. T. Sulaiman  
Department of Computer Engineering, College of Engineering,  
University of Mosul ,Iraq

Abstract. New integral inequality concerning triple integrals is presented.

## 1. Introduction

The following open question was proposed in [2]  
Under what conditions does the inequality

$$(1.1) \quad \int_0^1 f^{\alpha+\beta}(x) dx \geq \int_0^1 x^{\beta} f^{\alpha}(x) dx$$

hold for  $\alpha$  and  $\beta$  ?

In[1], the authors gave an answer by establishing the following

Theorem . *If the function  $f$  satisfies*

$$(1.2) \quad \int_0^1 f(t) dt \geq \frac{1-x^2}{2}, \quad \forall x \in [0,1],$$

then

$$\int_0^1 f^{\alpha+\beta}(x) dx \geq \int_0^1 x^{\beta} f^{\alpha}(x) dx$$

for every real  $\alpha \geq 1$  and  $\beta > 0$ .

The aim of this paper is that to give a new theorem concerning a similar result but for triple integrals and over the domain  $[a,b] \times [a,b] \times [a,b]$ .

## 2. New inequality

We state and prove the following

**Theorem 2.1.** *Let  $f(x,y)$  ,  $g(x)$  ,  $h(y)$  ,  $k(z)$  be continuous functions defined on  $[a,b] \times [a,b] \times$*

$[a, b], [a, b]$ , respectively,  $f$  is nonnegative,  $g(a) = h(a) = k(a) = 0$ ,  $g'(x), h'(y), k'(z) \geq 1$ . Let  $\alpha \geq 1, \beta, \gamma, \delta > 0$ . If

$$(2.1) \quad \int_a^b \int_a^b \int_a^b f(t, u, v) dt du dv \geq \int_a^b \int_a^b \int_a^b g(t) g'(t) h(u) h'(u) k(z) k'(z) dt du dv,$$

$$\forall x, y, z \in [a, b]$$

$$= \frac{1}{8} (g^2(b) - g^2(x)) (h^2(b) - h^2(y)) (k^2(b) - k^2(z)),$$

then

$$(2.2) \quad \int_a^b \int_a^b \int_a^b f^{\alpha+\beta+\gamma+\delta}(x, y, z) dx dy dz \geq \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz,$$

provided one of the following holds :

$$(i) \quad g(x), h(y), k(z) \geq 1 \quad \forall x, y, z \in [a, b].$$

$$(ii) \quad g^{\frac{\alpha(\gamma+\delta)}{\beta+\gamma+\delta}}(b) h^{\frac{\alpha(\beta+\delta)}{\beta+\gamma+\delta}}(b) k^{\frac{\alpha(\beta+\gamma)}{\beta+\gamma+\delta}}(b) \geq \frac{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)}{\left[ \beta \left( \frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta} \right) + 1 \right] \left[ \gamma \left( \frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta} \right) + 1 \right]} \times \frac{1}{\left[ \delta \left( \frac{\alpha+\beta+\delta+1}{\alpha+\beta+\delta} \right) + 1 \right]}.$$

$$(iii) \quad \max \left\{ \frac{g^{\alpha(\gamma+\delta)}(b)}{h^{\beta+\gamma+\delta}(b) k^{\beta+\gamma+\delta}(b)}, \frac{h^{\alpha(\beta+\delta)}(b)}{g^{\beta+\gamma+\delta}(b) k^{\beta+\gamma+\delta}(b)}, \frac{k^{\alpha(\beta+\gamma)}(b)}{g^{\beta+\gamma+\delta}(b) h^{\beta+\gamma+\delta}(b)} \right\} \leq \frac{(\alpha+\beta+\gamma+\delta+1)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)}.$$

**Proof .** By changing the order of the integration, we have

$$\begin{aligned} & \int_a^b \int_a^b \int_a^b \int_a^b \int_a^b \int_a^b f(t, u, v) g^{\beta-1}(x) g'(x) h^{\gamma-1}(y) h'(y) k^{\delta-1}(z) k'(z) dt du dv dx dy dz \\ &= \int_a^b \int_a^b \left( \int_a^b \int_a^b \left( \int_a^b f(t, u, v) g^{\beta-1}(x) g'(x) dt dx \right) h^{\gamma-1}(y) h'(y) du dy \right) k^{\delta-1}(z) k'(z) dv dz \\ &= \int_a^b \int_a^b \left( \int_a^b \int_a^b \left( \int_a^t f(t, u, v) dt \int_a^t g^{\beta-1}(x) g'(x) dx \right) h^{\gamma-1}(y) h'(y) du dy \right) k^{\delta-1}(z) k'(z) dv dz \\ &= \frac{1}{\beta} \int_a^b \int_a^b \left( \int_a^b \int_a^b \left( \int_a^b f(t, u, v) g^\beta(t) dt \right) h^{\gamma-1}(y) h'(y) du dy \right) k^{\delta-1}(z) k'(z) dv dz \\ &= \frac{1}{\beta} \int_a^b \int_a^b \left( \int_a^b g^\beta(t) dt \int_a^b f(t, u, v) h^{\gamma-1}(y) h'(y) du dy \right) k^{\delta-1}(z) k'(z) dv dz \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\beta} \int_a^b \int_z^b \left( \int_a^b g^\beta(t) dt \int_a^b f(t, u, v) du \int_a^u h^{\gamma-1}(y) h'(y) dy \right) k^{\delta-1}(z) k'(z) dv dz \\
&= \frac{1}{\beta \gamma} \int_a^b \int_z^b \left( \int_a^b g^\beta(t) dt \int_a^b f(t, u, v) h^\gamma(u) du \right) k^{\delta-1}(z) k'(z) dv dz \\
&= \frac{1}{\beta \gamma} \int_a^b \int_z^b \left( \int_a^b \int_a^b f(t, u, v) g^\beta(t) h^\gamma(u) dt du \right) k^{\delta-1}(z) k'(z) dv dz \\
&= \frac{1}{\beta \gamma} \int_a^b \int_a^b g^\beta(t) h^\gamma(u) dt du \int_a^b \int_z^b f(t, u, v) k^{\delta-1}(z) k'(z) dv dz \\
&= \frac{1}{\beta \gamma} \int_a^b \int_a^b g^\beta(t) h^\gamma(u) dt du \int_a^b f(t, u, v) dv \int_a^v k^{\delta-1}(z) k'(z) dz \\
&= \frac{1}{\beta \gamma \delta} \int_a^b \int_a^b g^\beta(t) h^\gamma(u) dt du \int_a^b f(t, u, v) k^\delta(v) dv \\
&= \frac{1}{\beta \gamma \delta} \int_a^b \int_a^b \int_a^b f(t, u, v) g^\beta(t) h^\gamma(u) k^\delta(v) dt du dv.
\end{aligned}$$

Also,

$$\begin{aligned}
&\int_a^b \int_a^b \int_a^b \int_a^b \int_a^b \int_a^b f(t, u, v) g^{\beta-1}(x) g'(x) h^{\gamma-1}(y) h'(y) k^{\delta-1}(z) k'(z) dt du dx dy dv dz \\
&= \int_a^b \int_a^b \left( \int_y^b \int_x^b f(t, u, v) dt du dv \right) g^{\beta-1}(x) g'(x) h^{\gamma-1}(y) h'(y) k^{\delta-1}(z) k'(z) dx dy dz \\
&\geq \frac{1}{8} \int_a^b \int_a^b \int_a^b \left( g^2(b) - g^2(x) \right) \left( h^2(b) - h^2(y) \right) \left( k^2(b) - k^2(z) \right) g^{\beta-1}(x) g'(x) h^{\gamma-1}(y) \\
&\quad \times h'(y) k^{\delta-1}(z) k'(z) dx dy dz \\
&= \frac{1}{8} \int_a^b \left( g^2(b) - g^2(x) \right) g^{\beta-1}(x) g'(x) dx \int_a^b \left( h^2(b) - h^2(y) \right) h^{\gamma-1}(y) h'(y) dy \\
&\quad \times \int_a^b \left( k^2(b) - k^2(z) \right) k^{\delta-1}(z) k'(z) dz \\
&= \frac{1}{8} \left( \frac{g^{\beta+2}(b)}{\beta} - \frac{g^{\beta+2}(b)}{\beta+1} \right) \left( \frac{h^{\gamma+2}(b)}{\gamma} - \frac{h^{\gamma+2}(b)}{\gamma+2} \right) \left( \frac{k^{\delta+2}(b)}{\delta} - \frac{k^{\delta+2}(b)}{\delta+2} \right) \\
&= \frac{g^{\beta+2}(b) h^{\gamma+2}(b) k^{\delta+2}(b)}{\beta \gamma \delta (\beta+2)(\gamma+2)(\delta+2)}.
\end{aligned}$$

Therefore

$$\int_a^b \int_a^b \int_a^b f(t, u, v) g^\beta(t) h^\gamma(u) k^\delta(v) dt du dv \geq \frac{g^{\beta+2}(b) h^{\gamma+2}(b) k^{\delta+2}(b)}{(\beta+2)(\gamma+2)(\delta+2)}.$$

Now, by the AG inequality,

$$\frac{1}{\alpha} f^{\alpha}(x, y, z) + \frac{\alpha-1}{\alpha} g^{\alpha}(x) h^{\alpha}(y) k^{\alpha}(z) \geq f(x, y, z) g^{\alpha-1}(x) h^{\alpha-1}(y) k^{\alpha-1}(z).$$

Multiplying the above inequality by  $\alpha g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z)$ , we obtain

$$\begin{aligned} f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z) &\geq \alpha f(x, y, z) g^{\alpha+\beta-1}(x) h^{\alpha+\gamma-1}(y) k^{\alpha+\delta-1}(z) \\ &\quad - (\alpha-1) g^{\alpha+\beta}(x) h^{\alpha+\gamma}(y) k^{\alpha+\delta}(z) \\ &\geq \alpha f(x, y, z) g^{\alpha+\beta-1}(x) h^{\alpha+\gamma-1}(y) k^{\alpha+\delta-1}(z) \\ &\quad - (\alpha-1) g^{\alpha+\beta}(x) g'(x) h^{\alpha+\gamma}(y) h'(y) k^{\alpha+\delta}(z) k'(z) \end{aligned}$$

Triple integration implies

$$\begin{aligned} \int_a^b \int_a^b \int_a^b f^{\alpha+\beta+\gamma+\delta}(x, y, z) dx dy dz &\geq \alpha \int_a^b \int_a^b \int_a^b f(x, y, z) g^{\alpha+\beta-1}(x) h^{\alpha+\gamma-1}(y) k^{\alpha+\delta-1}(z) dx dy dz \\ &\quad - (\alpha-1) \int_a^b \int_a^b \int_a^b g^{\alpha+\beta}(x) g'(x) h^{\alpha+\gamma}(y) h'(y) k^{\alpha+\delta}(z) k'(z) dx dy dz \\ &\geq \alpha \frac{g^{\alpha+\beta+1}(b) h^{\alpha+\gamma+1}(b) k^{\alpha+\delta+1}(b)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)} - (\alpha-1) \frac{g^{\alpha+\beta+1}(b) h^{\alpha+\gamma+1}(b) k^{\alpha+\delta+1}(b)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)} \\ &= \frac{g^{\alpha+\beta+1}(b) h^{\alpha+\gamma+1}(b) k^{\alpha+\delta+1}(b)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)}. \end{aligned}$$

We now deal with the three cases separately

**Case 1.** We claim that  $\beta \left( \frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta} \right) \leq \alpha+\beta$ ,  $\gamma \left( \frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta} \right) \leq \alpha+\gamma$ ,

and  $\delta \left( \frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta} \right) \leq \alpha+\delta$ . In fact

$$\beta \leq \beta + \gamma + \delta \Rightarrow 1 + \frac{\alpha}{\beta + \gamma + \delta} \leq 1 + \frac{\alpha}{\beta} \Rightarrow \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \leq \frac{\alpha + \beta}{\beta} \Rightarrow \beta \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right) \leq \alpha + \beta.$$

Therefore, we have

$$f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z) \leq \frac{\alpha}{\alpha + \beta + \gamma + \delta} f^{\alpha+\beta+\gamma+\delta}(x, y, z)$$



$$\begin{aligned}
(2.3) \quad & + \frac{\beta + \gamma + \delta}{\alpha + \beta + \gamma + \delta} \left( g^\beta(x) h^\gamma(y) k^\delta(z) \right)^{\frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta}} \\
& \leq \frac{\alpha}{\alpha + \beta + \gamma + \delta} f^{\alpha + \beta + \gamma + \delta}(x, y, z) + \frac{\beta + \gamma + \delta}{\alpha + \beta + \gamma + \delta} g^{\beta \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right)}(x) g'(x) \\
& \quad \times h^{\gamma \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right)}(y) h'(y) k^{\delta \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right)}(z) k'(z) \\
& \leq \frac{\alpha}{\alpha + \beta + \gamma + \delta} f^{\alpha + \beta + \gamma + \delta}(x, y, z) + \frac{\beta + \gamma + \delta}{\alpha + \beta + \gamma + \delta} g^{\alpha + \beta}(x) g'(x) h^{\alpha + \gamma}(y) h'(y) \\
& \quad \times k^{\alpha + \delta}(z) k'(z) .
\end{aligned}$$

Triple integrating both sides the above inequality from a to b gives

$$\begin{aligned}
& \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^{\alpha + \beta + \gamma + \delta}(x, y, z) dx dy dz \geq \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz \\
& \quad - \frac{\beta + \gamma + \delta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b g^{\alpha + \beta}(x) g'(x) h^{\alpha + \gamma}(y) h'(y) k^{\alpha + \delta}(z) k'(z) dx dy dz \\
& \geq \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz \\
& \quad + \frac{\beta + \gamma + \delta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) - g^{\alpha + \beta}(x) g'(x) h^{\alpha + \gamma}(y) h'(y) k^{\alpha + \delta}(z) k'(z) \right) \\
& \quad \times dx dy dz \\
& \geq \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz \\
& \quad + \frac{\alpha + \beta + \delta}{\alpha + \beta + \gamma + \delta} \left( \frac{g^{\alpha + \beta + 1}(b) h^{\alpha + \gamma + 1}(b) k^{\alpha + \delta + 1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} - \frac{g^{\alpha + \beta + 1}(a) h^{\alpha + \gamma + 1}(a) k^{\alpha + \delta + 1}(a)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} \right) \\
& = \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz .
\end{aligned}$$

**Case 2.** Since  $g(x) > 0$ ,  $g$  is increasing. Hence  $g(x) \geq g(a) = 0$ . Also  $h(u), k(v) > 0$ . It is not difficult to show that (ii) implies

$$\frac{g^{\beta \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right) + 1}(b) h^{\gamma \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right) + 1}(b) k^{\delta \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right) + 1}(b)}{\left[ \beta \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right) + 1 \right] \left[ \gamma \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right) + 1 \right] \left[ \delta \left( \frac{\alpha + \beta + \gamma + \delta}{\beta + \gamma + \delta} \right) + 1 \right]} .$$

$$\leq \frac{g^{\alpha+\beta+1}(b)h^{\alpha+\gamma+1}(b)k^{\alpha+\delta+1}(b)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)}.$$

Therefore, from (2.3) we have

$$\begin{aligned} & \frac{\alpha}{\alpha+\beta+\gamma+\delta} \int_a^b \int_a^b \int_a^b f^{\alpha+\beta+\gamma+\delta}(x, y, z) dx dy dz \\ & \geq \frac{\alpha}{\alpha+\beta+\gamma+\delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz \\ & \quad + \frac{\beta+\gamma+\delta}{\alpha+\beta+\gamma+\delta} \left( \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz \right. \\ & \quad \left. - \int_a^b \int_a^b \int_a^b g^{\beta\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)}(x) g'(x) h^{\gamma\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)}(y) h'(y) k^{\delta\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)}(z) k'(z) dx dy dz \right) \\ & \geq \frac{\alpha}{\alpha+\beta+\gamma+\delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz + \frac{\beta+\gamma+\delta}{\alpha+\beta+\gamma+\delta} \times \\ & \quad \left( \frac{g^{\alpha+\beta+1}(b)h^{\alpha+\gamma+1}(b)k^{\alpha+\delta+1}(b)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)} \right. \\ & \quad \left. - \frac{g^{\beta\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)}(b)h^{\gamma\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)}(b)k^{\delta\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)}(b)}{\left[\beta\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)+1\right]\left[\gamma\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)+1\right]\left[\delta\left(\frac{\alpha+\beta+\gamma+\delta}{\beta+\gamma+\delta}\right)+1\right]} \right) \\ & \geq \frac{\alpha}{\alpha+\beta+\gamma+\delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz + \frac{\beta+\gamma+\delta}{\alpha+\beta+\gamma+\delta} \times \\ & \quad \left( \frac{g^{\alpha+\beta+1}(b)h^{\alpha+\gamma+1}(b)k^{\alpha+\delta+1}(b)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)} - \frac{g^{\alpha+\beta+1}(b)h^{\alpha+\gamma+1}(b)k^{\alpha+\delta+1}(b)}{(\alpha+\beta+1)(\alpha+\gamma+1)(\alpha+\delta+1)} \right) \\ & = \frac{\alpha}{\alpha+\beta+\gamma+\delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz. \end{aligned}$$

**Case 3.** We have, by (iii),

$$f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) \leq \frac{\alpha}{\alpha+\beta+\gamma+\delta} f^{\alpha+\beta+\gamma+\delta}(x, y, z) + \frac{\beta}{\alpha+\beta+\gamma+\delta} g^{\alpha+\beta+\gamma+\delta}(x)$$

$$\begin{aligned}
& + \frac{\gamma}{\alpha + \beta + \gamma + \delta} h^{\alpha + \beta + \gamma + \delta}(y) + \frac{\delta}{\alpha + \beta + \gamma + \delta} k^{\alpha + \beta + \gamma + \delta}(z) \\
& \leq \frac{\alpha}{\alpha + \beta + \gamma + \delta} f^{\alpha + \beta + \gamma + \delta}(x, y, z) + \frac{\beta}{\alpha + \beta + \gamma + \delta} g^{\alpha + \beta + \gamma + \delta}(x) g'(x) \\
& \quad + \frac{\gamma}{\alpha + \beta + \gamma + \delta} h^{\alpha + \beta + \gamma + \delta}(y) h'(y) + \frac{\delta}{\alpha + \beta + \gamma + \delta} k^{\alpha + \beta + \gamma + \delta}(z) k'(z),
\end{aligned}$$

which implies

$$\begin{aligned}
& \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^{\alpha + \beta + \gamma + \delta}(x, y, z) dx dy dz \geq \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) \\
& \quad \times k^{\delta}(z) dx dy dz \\
& \quad + \frac{\beta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z) - g^{\alpha + \beta + \gamma + \delta}(x) g'(x) \right) dx dy dz \\
& \quad + \frac{\gamma}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z) - h^{\alpha + \beta + \gamma + \delta}(y) h'(y) \right) dx dy dz \\
& \quad + \frac{\delta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z) - k^{\alpha + \beta + \gamma + \delta}(z) k'(z) \right) dx dy dz \\
& \geq \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z) dx dy dz \\
& \quad + \frac{\beta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( \frac{g^{\alpha + \beta + 1}(b) h^{\alpha + \gamma + 1}(b) k^{\alpha + \delta + 1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} - \frac{g^{\alpha + \beta + \gamma + \delta + 1}(b)(b - a)^2}{\alpha + \beta + \gamma + \delta + 1} \right) \\
& \quad + \frac{\gamma}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( \frac{g^{\alpha + \beta + 1}(b) h^{\alpha + \gamma + 1}(b) k^{\alpha + \delta + 1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} - \frac{h^{\alpha + \beta + \gamma + \delta + 1}(b)(b - a)^2}{\alpha + \beta + \gamma + \delta + 1} \right) \\
& \quad + \frac{\delta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( \frac{g^{\alpha + \beta + 1}(b) h^{\alpha + \gamma + 1}(b) k^{\alpha + \delta + 1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} - \frac{k^{\alpha + \beta + \gamma + \delta + 1}(b)(b - a)^2}{\alpha + \beta + \gamma + \delta + 1} \right) \\
& \geq \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^{\alpha}(x, y, z) g^{\beta}(x) h^{\gamma}(y) k^{\delta}(z) dx dy dz \\
& \quad + \frac{\beta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( \frac{g^{\alpha + \beta + 1}(b) h^{\alpha + \gamma + 1}(b) k^{\alpha + \delta + 1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} - \frac{g^{\alpha + \beta + 1}(b) h^{\alpha + \gamma + 1}(b) k^{\alpha + \delta + 1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} \right)
\end{aligned}$$

$$\begin{aligned}
& + \frac{\gamma}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( \frac{g^{\alpha+\beta+1}(b) h^{\alpha+\gamma+1}(b) k^{\alpha+\delta+1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} - \frac{g^{\alpha+\beta+1}(b) h^{\alpha+\gamma+1}(b) k^{\alpha+\delta+1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} \right) \\
& + \frac{\delta}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b \left( \frac{g^{\alpha+\beta+1}(b) h^{\alpha+\gamma+1}(b) k^{\alpha+\delta+1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} - \frac{g^{\alpha+\beta+1}(b) h^{\alpha+\gamma+1}(b) k^{\alpha+\delta+1}(b)}{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)} \right) \\
& = \frac{\alpha}{\alpha + \beta + \gamma + \delta} \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) g^\beta(x) h^\gamma(y) k^\delta(z) dx dy dz.
\end{aligned}$$

### 3. Applications

**Corollary 3.1.** Suppose that the statement of Theorem 2.1 is satisfied. If

$$(3.1) \quad \int_a^b \int_a^b \int_a^b f(t, u, v) dt du dv \geq \int_a^b \int_a^b \int_a^b t u v dt du dv \quad \forall x, y, z \in [0, b],$$

then

$$(3.2) \quad \int_0^b \int_0^b \int_0^b f^{\alpha+\beta+\gamma+\delta}(x, y, z) dx dy dz \geq \int_0^b \int_0^b \int_0^b f^\alpha(x, y, z) x^\beta y^\gamma z^\delta dx dy dz,$$

provided

$$b^\alpha \geq \frac{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)}{\left[ \beta \left( \frac{\alpha}{\beta + \gamma + \delta} \right) + 1 \right] \left[ \gamma \left( \frac{\alpha}{\beta + \gamma + \delta} \right) + 1 \right] \left[ \delta \left( \frac{\alpha}{\beta + \gamma + \delta} \right) + 1 \right]}.$$

**Proof .** Follows from Theorem 2.1, using (ii) and putting  $a = 0$ ,  $g(z) = h(z) = k(z) = z$ .

**Corollary 3.2.** Suppose that the statement of Theorem 2.1 is satisfied . If

$$(3.3) \quad \int_a^b \int_a^b \int_a^b f(t, u, v) dt du dv \geq \int_a^b \int_a^b \int_a^b (t - a)(u - a)(v - a) dt du dv \quad \forall x, y, z \in [a, b],$$

then

$$(3.4) \quad \int_a^b \int_a^b \int_a^b f^{\alpha+\beta+\gamma+\delta}(x, y, z) dx dy dz \geq \int_a^b \int_a^b \int_a^b f^\alpha(x, y, z) (x - a)^\beta (y - a)^\gamma (z - a)^\delta dx dy dz,$$

provided

$$(b - a)^\alpha \geq \frac{(\alpha + \beta + 1)(\alpha + \gamma + 1)(\alpha + \delta + 1)}{(\alpha + \beta + \gamma + \delta + 1)}.$$

**Proof .** Follows from Theorem 2.1, by putting  $g(z) = h(z) = k(z) = z - a$ .

## References

- [1] K. BOUKERRIOUA and A. GUEZANE\_LAKOUD, On an open question Regarding an integral inequality, J. Ineq. Pure and Appl. Math, 8(3) (2007), Art. 77.
- [2] QUOC ANH NGO and DU DUC THANG, TRAN TAT DAT and DANG ANH THAN, Notes on an integral inequality, J. Ineq. Pure and Appl. Math., 7(4) (2006) Art. 120.

# On some inequalities for Concave Functions

W. T. Sulaiman

Department of Computer Engineering, College  
of Engineering, University of Mosul.

**Abstract.** Several new inequalities concerning improvement, generalization and new results of concave functions are obtained .

## 1. Introduction

A real valued function  $f$  defined on an interval  $I \subset \mathbb{R}$  will be termed convex if

$$f(ta + (1-t)b) \leq tf(a) + (1-t)f(b), \quad a, b \in I, \quad 0 \leq t \leq 1,$$

or equivalently

$$f\left(\frac{a+b}{2}\right) \leq \frac{f(a) + f(b)}{2}.$$

If the above inequalities are reversed then  $f$  is said to be concave .

In [1], Pearce and Pecaric established the following results concerning concave functions

**Theorem 1.1.** *If  $f : [a, b] \rightarrow \mathbb{R}^+$  is continuous and concave, then for all real numbers  $\alpha, \beta$  with  $0 < \alpha + \beta \leq 1, 0 < \beta \leq 1$ ,*

$$\begin{aligned} \frac{1}{(b-a)^2} \int_a^b f^\alpha(t) dt \int_a^b f^\beta(t) dt &\leq \frac{\alpha + 2\beta}{\alpha + \beta} f^{\alpha+\beta}\left(\frac{a+b}{2}\right) \\ &\quad - \frac{\beta}{\alpha + \beta} \frac{f^{\alpha+\beta}(a) + f^{\alpha+\beta}(b)}{2}. \end{aligned}$$

**Theorem 1.2.** *If  $f : [a, b] \rightarrow \mathbb{R}^+$  is continuous and concave, then for all real numbers  $\beta$  with  $0 < \beta \leq 1$ ,*

$$\frac{1}{(b-a)^2} \int_a^b f^\beta(t) dt \int_a^b f^{-\beta}(t) dt \leq 1 + \beta \log \left( \frac{f((a+b)/2)}{\sqrt{f(a)f(b)}} \right).$$

## 2. Results

**Lemma 2.1.** Let  $a, b \geq 0$ . Then

$$(2.1) \quad \left( \frac{a+b}{2} \right)^\alpha \geq \frac{a^\alpha + b^\alpha}{2}, \quad 0 \leq \alpha \leq 1,$$

$$(2.2) \quad \left( \frac{a+b}{2} \right)^\alpha \leq \frac{a^\alpha + b^\alpha}{2}, \quad \alpha \geq 1.$$

**Proof.** Let  $0 \leq x \leq 1$ ,  $0 \leq \alpha \leq 1$ . Consider

$$f(x) = \left( \frac{x+1}{2} \right)^\alpha - \frac{x^\alpha + 1}{2}.$$

We have

$$f'(x) = \frac{\alpha}{2} \left( \left( \frac{x+1}{2} \right)^{\alpha-1} - x^{\alpha-1} \right) \geq 0.$$

Therefore  $f$  is non-decreasing. Since  $f(0) = \frac{1}{2^\alpha} - \frac{1}{2} \geq 0$ , then, we have

$$f(x) \geq f(0) \geq 0.$$

Set  $c = \min\{a, b\}$ ,  $C = \max\{a, b\}$ , then (2.1) follows by putting  $x = c/C$ . The proof of the other part is similar.

**Remark 1.** The above Lemma shows the following

- (i)  $f \geq 0$ ,  $f$  concave  $\Rightarrow f^\alpha$  concave if  $0 \leq \alpha \leq 1$ .
- (ii)  $f \geq 0$ ,  $f$  convex  $\Rightarrow f^\alpha$  convex if  $\alpha \geq 1$ .

$$(i) \text{ follows as } f^\alpha \left( \frac{a+b}{2} \right) \geq \left( \frac{f(a) + f(b)}{2} \right)^\alpha \geq \frac{f^\alpha(a) + f^\alpha(b)}{2}.$$

The proof of the other part is the same.

**Lemma 2.2.** Let  $a, b > 0$ . then

$$(2.3) \quad \left( \frac{a+b}{2} \right)^{-1} \leq \frac{a^{-1} + b^{-1}}{2}.$$

**Proof.** Since  $(b-a)^2 \geq 0 \Rightarrow 2ab \leq a^2 + b^2 \Rightarrow \frac{a}{b} + \frac{b}{a} \geq 2$ , then, we have

$$(a+b)(a^{-1} + b^{-1}) = 2 + \frac{a}{b} + \frac{b}{a} \geq 4,$$

which implies (2.3).

**Remark 2.** If  $f^{-1}(a) = f(a)^{-1}$ , then

$$f \geq 0, f \text{ concave} \Rightarrow f^{-1} \text{ convex}.$$

This follows from Lemma 2.2 :

$$f \text{ concave implies } f^{-1} \left( \frac{a+b}{2} \right) \leq \left( \frac{f(a) + f(b)}{2} \right)^{-1} \leq \frac{f^{-1}(a) + f^{-1}(b)}{2}.$$

The following giving a refinement for the generalization of Hadamard's inequality.

**Theorem 2.2.** Let  $f : [a, b] \rightarrow \mathfrak{R}$  be convex, and  $g : [a, b] \rightarrow \mathfrak{R}_0^+$  be integrable and symmetric with respect to the lines  $\frac{3a+b}{4}, \frac{a+3b}{4}$ . Then

$$(2.4) \quad f\left(\frac{a+b}{2}\right) \int_a^b g(x) dx \leq \int_a^b f(x) g(x) dx \leq \frac{1}{2} \left( f(a) + f\left(\frac{a+b}{2}\right) \right) \int_a^{\frac{a+b}{2}} g(x) dx \\ + \frac{1}{2} \left( f\left(\frac{a+b}{2}\right) + f(b) \right) \int_{\frac{a+b}{2}}^b g(x) dx.$$

In particular for  $g(x) = 1$ ,

$$(2.5) \quad f\left(\frac{a+b}{2}\right) \leq \frac{1}{b-a} \int_a^b f(x) dx \leq \frac{1}{2} f\left(\frac{a+b}{2}\right) + \frac{f(a) + f(b)}{4} \leq \frac{f(a) + f(b)}{2}.$$

If  $f$  is concave, (2.4), (2.5) reversed.

**Proof.** It is sufficient to prove the middle inequality. As  $f$  convex implies

$$\begin{aligned} f\left(a + \frac{a+b}{2} - x\right) &= f\left(a + \frac{a+b}{2} - ta - (1-t)\frac{a+b}{2}\right), \quad 0 \leq t \leq 1 \\ &= f\left((1-t)a + t\frac{a+b}{2}\right) \\ &\leq (1-t)f(a) + tf\left(\frac{a+b}{2}\right) \\ &= f(a) + f\left(\frac{a+b}{2}\right) - \left(tf(a) + (1-t)f\left(\frac{a+b}{2}\right)\right) \\ &\leq f(a) + f\left(\frac{a+b}{2}\right) - f\left(ta + (1-t)\frac{a+b}{2}\right) \\ &= f(a) + f\left(\frac{a+b}{2}\right) - f(x), \end{aligned}$$

then, we have

$$\begin{aligned} \int_a^b f(x) g(x) dx &= \int_a^{\frac{a+b}{2}} f(x) g(x) dx + \int_{\frac{a+b}{2}}^b f(x) g(x) dx = I_1 + I_2, \\ I_1 &= \int_a^{\frac{a+b}{2}} f(x) g(x) dx \leq \int_a^{\frac{a+b}{2}} \left( f(a) + f\left(\frac{a+b}{2}\right) - f\left(a + \frac{a+b}{2} - x\right) \right) g(x) dx \\ &= \left( f(a) + f\left(\frac{a+b}{2}\right) \right) \int_a^{\frac{a+b}{2}} g(x) dx - \int_a^{\frac{a+b}{2}} f\left(a + \frac{a+b}{2} - x\right) g(x) dx \end{aligned}$$



$$\begin{aligned}
&= \left( f(a) + f\left(\frac{a+b}{2}\right) \right) \int_a^{\frac{a+b}{2}} g(x) dx - \int_a^{\frac{a+b}{2}} f\left(a + \frac{a+b}{2} - x\right) g\left(a + \frac{a+b}{2} - x\right) dx \\
&= \left( f(a) + f\left(\frac{a+b}{2}\right) \right) \int_a^{\frac{a+b}{2}} g(x) dx - \int_a^{\frac{a+b}{2}} f(x) g(x) dx,
\end{aligned}$$

which implies

$$I_1 \leq \frac{1}{2} \left( f(a) + f\left(\frac{a+b}{2}\right) \right) \int_a^{\frac{a+b}{2}} g(x) dx.$$

Similarly,

$$I_2 \leq \frac{1}{2} \left( f\left(\frac{a+b}{2}\right) + f(b) \right) \int_{\frac{a+b}{2}}^b g(x) dx.$$

The following is an improvement of Theorems 1.1 and 1.2 via simpler method

**Theorem 2.3.** If  $f : [a, b] \rightarrow \mathbb{R}^+$  is continuous and concave, then for all real numbers  $\alpha, \beta$  with  $0 < \alpha + \beta \leq 1, 0 < \beta \leq 1$ ,

$$\begin{aligned}
(2.6) \quad \frac{1}{(b-a)^2} \int_a^b f^\alpha(t) dt \int_a^b f^\beta(t) dt &\leq \frac{\alpha + \frac{3}{2}\beta}{\alpha + \beta} f^{\alpha+\beta}\left(\frac{a+b}{2}\right) \\
&\quad - \frac{\beta}{\alpha + \beta} \frac{f^{\alpha+\beta}(a) + f^{\alpha+\beta}(b)}{4}
\end{aligned}$$

**Proof.** On the light of inequality (2.5) for concave functions

$$\begin{aligned}
\frac{1}{(b-a)^2} \int_a^b f^\alpha(t) dt \int_a^b f^\beta(t) dt &= \frac{1}{(b-a)^2} \int_a^b \int_a^b f^\alpha(x) f^\beta(t) dx dt \\
&\leq \frac{1}{(b-a)^2} \int_a^b \int_a^b \left( \frac{\alpha}{\alpha + \beta} f^{\alpha+\beta}(x) + \frac{\beta}{\alpha + \beta} f^{\alpha+\beta}(t) \right) dx dt \\
&= \frac{1}{b-a} \int_a^b \left( \frac{\alpha}{\alpha + \beta} f^{\alpha+\beta}(x) dx + \frac{\beta}{\alpha + \beta} f^{\alpha+\beta}(t) dt \right) \\
&= \frac{1}{b-a} \int_a^b f^{\alpha+\beta}(t) dt \\
&= \frac{\alpha + 2\beta}{\alpha + \beta} \frac{1}{b-a} \int_a^b f^{\alpha+\beta}(t) dt - \frac{\beta}{\alpha + \beta} \frac{1}{b-a} \int_a^b f^{\alpha+\beta}(t) dt \\
&\leq \frac{\alpha + 2\beta}{\alpha + \beta} f^{\alpha+\beta}\left(\frac{a+b}{2}\right) - \frac{\beta}{\alpha + \beta} \\
&\quad \times \left( \frac{1}{2} f^{\alpha+\beta}\left(\frac{a+b}{2}\right) + \frac{f^{\alpha+\beta}(a) + f^{\alpha+\beta}(b)}{4} \right)
\end{aligned}$$

$$= \frac{\alpha + \frac{3}{2}\beta}{\alpha + \beta} f^{\alpha+\beta} \left( \frac{a+b}{2} \right) - \frac{\beta}{\alpha + \beta} \frac{f^{\alpha+\beta}(a) + f^{\alpha+\beta}(b)}{4}.$$

**Theorem 2.4.** If  $f : [a, b] \rightarrow \mathfrak{R}^+$  is numbers  $\beta$  with  $0 < \beta \leq 1$ , continuous and concave, then for all real

$$(2.7) \quad 1 \leq \frac{1}{(b-a)^2} \int_a^b f^\beta(t) dt \int_a^b f^{-\beta}(t) dt \leq 1 + \frac{1}{2} \beta \ln \left( \frac{f((a+b)/2)}{\sqrt{f(a)f(b)}} \right)$$

**Proof.**

$$1 = \frac{1}{b-a} \int_a^b f^{\beta/2}(t) f^{-\beta/2}(t) dt \leq \frac{1}{b-a} \left( \int_a^b f^\beta(t) dt \right)^{1/2} \left( \int_a^b f^{-\beta}(t) dt \right)^{1/2},$$

Squaring the above implies the left inequality.

The right inequality in (2.6) can be written in the form

$$\begin{aligned} \frac{1}{(b-a)^2} \int_a^b f^\alpha(t) dt \int_a^b f^\beta(t) dt \leq 1 + \left( \alpha + \frac{3}{2}\beta \right) \frac{f^{\alpha+\beta} \left( \frac{a+b}{2} \right) - 1}{\alpha + \beta} \\ - \frac{\beta}{4} \left( \frac{f^{\alpha+\beta}(a) - 1}{\alpha + \beta} + \frac{f^{\alpha+\beta}(b) - 1}{\alpha + \beta} \right). \end{aligned}$$

Now making use of the limit,  $\lim_{t \rightarrow 0} \frac{a^t - 1}{t} = \ln a$ , the above inequality gives the right inequality in (2.7) as  $\alpha \rightarrow -\beta$ .

The following result giving the reverse case

**Theorem 2.5.** If  $f : [a, b] \rightarrow \mathfrak{R}^+$  is continuous and concave, then for all real numbers  $\beta > 0$ ,  $0 < \alpha - \beta \leq 1$ ,

$$(2.8) \quad \frac{1}{(b-a)^2} \int_a^b f^\alpha(t) dt \int_a^b f^{-\beta}(t) dt \geq \frac{\frac{1}{2}\alpha - \beta}{\alpha - \beta} f^{\alpha-\beta} \left( \frac{a+b}{2} \right) \\ + \frac{\alpha}{\alpha - \beta} \frac{f^{\alpha-\beta}(a) + f^{\alpha-\beta}(b)}{4}.$$

In particular

$$(2.9) \quad \frac{1}{(b-a)^2} \int_a^b f^\beta(t) dt \int_a^b f^{-\beta}(t) dt \geq 1 + \beta \ln \left( \frac{f(a)f(b)}{\sqrt{f((a+b)/2)}} \right).$$

**Proof.**

$$\frac{1}{(b-a)^2} \int_a^b f^\alpha(t) dt \int_a^b f^{-\beta}(t) dt \geq \frac{1}{(b-a)^2} \int_a^b \int_a^b \left( \frac{\alpha}{\alpha - \beta} f^{\alpha-\beta}(x) - \frac{\beta}{\alpha - \beta} f^{\alpha-\beta}(t) \right) dx dt$$

$$\begin{aligned}
&= \frac{1}{(b-a)} \int_a^b f^{\alpha-\beta}(t) dt \\
&= \frac{\alpha}{\alpha-\beta} \int_a^b f^{\alpha-\beta}(t) dt - \frac{\beta}{\alpha-\beta} \int_a^b f^{\alpha-\beta}(t) dt \\
&\geq \frac{\alpha}{\alpha-\beta} \left( \frac{1}{2} f^{\alpha-\beta} \left( \frac{a+b}{2} \right) + \frac{f^{\alpha-\beta}(a) + f^{\alpha-\beta}(b)}{4} \right) \\
&\quad - \frac{\beta}{\alpha-\beta} f^{\alpha-\beta} \left( \frac{a+b}{2} \right) \\
&= \frac{\frac{1}{2}\alpha - \beta}{\alpha-\beta} f^{\alpha-\beta} \left( \frac{a+b}{2} \right) + \frac{\alpha}{\alpha-\beta} \frac{f^{\alpha-\beta}(a) + f^{\alpha-\beta}(b)}{4}.
\end{aligned}$$

Inequality (2.8) can be written as

$$\begin{aligned}
\frac{1}{(b-a)^2} \int_a^b f^{\alpha}(t) dt \int_a^b f^{-\beta}(t) dt &\geq 1 + \left( \frac{1}{2}\alpha - \beta \right) \frac{f^{\alpha-\beta} \left( \frac{a+b}{2} \right) - 1}{\alpha - \beta} \\
&\quad + \frac{\alpha}{4} \left( \frac{f(a)}{\alpha - \beta} + \frac{f^{\alpha-\beta}(b) - 1}{\alpha - \beta} \right).
\end{aligned}$$

By letting  $\alpha \rightarrow \beta$ , the above implies

$$\frac{1}{(b-a)^2} \int_a^b f^{\beta}(t) dt \int_a^b f^{-\beta}(t) dt \geq 1 + \beta \ln \frac{f(a)f(b)}{\sqrt{f((a+b)/2)}}.$$

Other kinds are also obtained

**Theorem 2.6.** If  $f : [a, b] \rightarrow \mathfrak{R}^+$  is continuous and concave, then for all real numbers  $\alpha \geq 0, \beta > 0, \alpha + \beta = 1$ ,

$$\begin{aligned}
(2.10) \quad \frac{1}{(b-a)^2} \int_a^b f^{\alpha}(t) dt \int_a^b f^{-\beta}(t) dt &\leq \alpha f \left( \frac{a+b}{2} \right) \\
&\quad + \beta \left( \frac{1}{2} f^{-1} \left( \frac{a+b}{2} \right) + \frac{f^{-1}(a) + f^{-1}(b)}{4} \right).
\end{aligned}$$

**Proof.** By virtue of Remark 2,

$$\begin{aligned}
\frac{1}{(b-a)^2} \int_a^b f^{\alpha}(t) dt \int_a^b f^{-\beta}(t) dt &= \frac{1}{(b-a)^2} \int_a^b \int_a^b f^{\alpha}(x) f^{-\beta}(t) dx dt \\
&\leq \frac{1}{(b-a)^2} \int_a^b \int_a^b \left( \frac{\alpha}{\alpha+\beta} f^{\alpha+\beta}(x) + \frac{\beta}{\alpha+\beta} f^{-(\alpha+\beta)}(t) \right) dx dt \\
&= \frac{1}{(b-a)^2} \int_a^b \int_a^b (\alpha f(x) + \beta f^{-1}(t)) dx dt
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{b-a} \int_a^b (\alpha f(x) dx + \beta f^{-1}(t) dt) \\
&\leq \alpha f\left(\frac{a+b}{2}\right) + \beta \left(\frac{1}{2} f^{-1}\left(\frac{a+b}{2}\right) + \frac{f^{-1}(a) + f^{-1}(b)}{4}\right).
\end{aligned}$$

**Theorem 2.7.** If  $f : [a, b] \rightarrow \mathfrak{R}^+$  is continuous and concave, then for all real numbers

$$\alpha, \beta, s, t, p, q, \quad p \leq 1/\alpha, \quad q \geq 1/\beta, \quad s < 1/\alpha, \quad t > -1/\beta, \quad \frac{1}{p} + \frac{1}{q} = 1, \quad \frac{1}{s} + \frac{1}{t} = 1, \quad p > 1, \quad 0 < s < 1,$$

$$\begin{aligned}
(2.11) \quad & \frac{1}{2s} \left( f^{\alpha s} \left( \frac{a+b}{2} \right) + \frac{f^{\alpha s}(a) + f^{\alpha s}(b)}{2} \right) + \frac{1}{2t} \left( f^{-\beta t} \left( \frac{a+b}{2} \right) + \frac{f^{-\beta t}(a) + f^{-\beta t}(b)}{2} \right) \\
& \leq \frac{1}{(b-a)^2} \int_a^b f^{\alpha}(x) dx \int_a^b f^{-\beta}(x) dx \leq \frac{1}{p} f^{\alpha p} \left( \frac{a+b}{2} \right) \\
& \quad + \frac{1}{2q} \left( f^{-\beta q} \left( \frac{a+b}{2} \right) + \frac{f^{-\beta q}(a) + f^{-\beta q}(b)}{2} \right).
\end{aligned}$$

**Proof.** We have

$$\begin{aligned}
& \frac{1}{(b-a)^2} \int_a^b f^{\alpha}(x) dx \int_a^b f^{-\beta}(u) du \leq \frac{1}{(b-a)^2} \int_a^b \int_a^b \left( \frac{1}{p} f^{\alpha p}(x) + \frac{1}{q} f^{-\beta q}(u) \right) dx du \\
& = \frac{1}{b-a} \int_a^b \left( \frac{1}{p} f^{\alpha p}(x) dx + \frac{1}{q} f^{-\beta q}(u) du \right) \\
& \leq \frac{1}{p} f^{\alpha p} \left( \frac{a+b}{2} \right) + \frac{1}{2q} \left( f^{-\beta q} \left( \frac{a+b}{2} \right) + \frac{f^{-\beta q}(a) + f^{-\beta q}(b)}{2} \right).
\end{aligned}$$

Also

$$\begin{aligned}
& \frac{1}{(b-a)^2} \int_a^b f^{\alpha}(x) dx \int_a^b f^{-\beta}(u) du \geq \frac{1}{(b-a)^2} \left( \frac{1}{s} f^{\alpha s}(x) + \frac{1}{t} f^{-\beta t}(u) \right) dx du \\
& = \frac{1}{b-a} \left( \frac{1}{s} f^{\alpha s}(x) dx + \frac{1}{t} f^{-\beta t}(u) du \right) \\
& \geq \frac{1}{2s} \left( f^{\alpha s} \left( \frac{a+b}{2} \right) + \frac{f^{\alpha s}(a) + f^{\alpha s}(b)}{2} \right) \\
& \quad + \frac{1}{2t} \left( f^{-\beta t} \left( \frac{a+b}{2} \right) + \frac{f^{-\beta t}(a) + f^{-\beta t}(b)}{2} \right).
\end{aligned}$$

**Theorem 2.8.** If  $f : [a, b] \rightarrow \mathfrak{R}^+$  is continuous and concave, then for all real numbers  $\alpha \geq 0, \beta > 0, \alpha + \beta \geq 1$ ,

$$\begin{aligned}
 (2.12) \quad & \frac{1}{(b-a)^2} \int_a^b f^{-\alpha}(x) dx \int_a^b f^{-\beta}(t) dt \\
 & \leq \frac{1}{2} \frac{\alpha}{\alpha + \beta} f^{-\alpha-\beta} \left( \frac{a+b}{2} \right) + \frac{\alpha + 2\beta}{\alpha + \beta} \frac{f^{-\alpha-\beta}(a) + f^{-\alpha-\beta}(b)}{4}
 \end{aligned}$$

**Proof.** Let  $g = f^{-1}$ . Then  $g$  is convex, and hence

$$\begin{aligned}
 & \frac{1}{(b-a)^2} \int_a^b g^{\alpha}(x) dx \int_a^b g^{\beta}(t) dt \leq \frac{1}{(b-a)} \left( \frac{\alpha}{\alpha + \beta} g^{\alpha+\beta}(x) dx + \frac{\beta}{\alpha + \beta} g^{\alpha+\beta}(t) dt \right) \\
 & = \frac{\alpha + 2\beta - \beta}{\alpha + \beta} \frac{1}{b-a} \int_a^b g^{\alpha+\beta}(t) dt \\
 & \leq \frac{\alpha + 2\beta}{\alpha + \beta} \left( \frac{1}{2} g^{\alpha+\beta} \left( \frac{a+b}{2} \right) + \frac{g^{\alpha+\beta}(a) + g^{\alpha+\beta}(b)}{4} \right) \\
 & \quad - \frac{\beta}{\alpha + \beta} g^{\alpha+\beta} \left( \frac{a+b}{2} \right) \\
 & = \frac{1}{2} \frac{\alpha}{\alpha + \beta} g^{\alpha+\beta} \left( \frac{a+b}{2} \right) + \frac{\alpha + 2\beta}{\alpha + \beta} \frac{g^{\alpha+\beta}(a) + g^{\alpha+\beta}(b)}{4}
 \end{aligned}$$

Therefore

$$\begin{aligned}
 & \frac{1}{(b-a)^2} \int_a^b f^{-\alpha}(x) dx \int_a^b f^{-\beta}(t) dt \\
 & \leq \frac{1}{2} \frac{\alpha}{\alpha + \beta} f^{-\alpha-\beta} \left( \frac{a+b}{2} \right) + \frac{\alpha + 2\beta}{\alpha + \beta} \frac{f^{-\alpha-\beta}(a) + f^{-\alpha-\beta}(b)}{4}.
 \end{aligned}$$

## References

- [1]. C. E. M. Pearce and J. E. Pecaric, On some inequalities of Brenner and Alzer for concave functions, J. Math. Anal. Appl. 198 (1996), 282-288.

## RECURRENCE RELATION WITH BINOMIAL COEFFICIENT

GEORGE GROSSMAN  
DEPARTMENT OF MATHEMATICS  
CENTRAL MICHIGAN UNIVERSITY  
MT. PLEASANT, MI 48859  
GROSS1GW@CMICH.EDU  
PHONE: 989-774-5577 FAX: 989-774-2414

AKLILU ZELEKE  
DEPARTMENT OF MATHEMATICS  
LYMAN BRIGGS COLLEGE  
EAST LANSING, MI 48825  
ZELEKE@MSU.EDU

XINYUN ZHU  
DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF TEXAS OF THE PERMIAN BASIN  
ODESSA, TX 79762  
ZHU\_X@UTPB.EDU

## ABSTRACT

We derive a linear, nonhomogeneous, recurrence relation having two indices of recurrence, with initial conditions and equated to a binomial coefficient. We relate a known classical combinatorial identity to partial sums of geometric series and show that a certain corresponding sum always is a rational, non-integer term. We construct solutions which are rational expressions with an indeterminate form evaluated in a limit as a binomial coefficient. We establish several interesting combinatorial identities and use these to express sums of powers of integers (in particular, squares and cubes) as a finite sum of binomial coefficients with integer coefficients.

Keywords: recurrence relation, binomial coefficient, characteristic polynomial.

## 1. Introduction

Different counting procedures and various arrangements of mathematical objects lend themselves to combinatorial identities. Moreover, combinatorial identities [11] arise in numerous settings in mathematics and commonly involve polynomials, binomial coefficients and recurrence relations and the current paper combines all of these features. For example, if  $S_p(n) = 1 + 2^p + \cdots + n^p$ , integers  $p > 0, n \geq 0$ , then it can be shown [9], (sect. 4 contains result) that  $S_p(n)$  satisfies a recurrence relation involving binomial coefficients, term  $1/(p+1)$  and  $S_i(n), 0 \leq i < n, S_0(n) = n$ . Additionally, computer-generated proofs [12] also generate identities as a by-product of the proof-process.

In a sense, the present paper offers an algorithmic way of producing at the least a sequence of combinatorial identities whose importance or significance lies in originality. In one case an elegant result [10] (see sect. 2) was generalized.

The basic idea lies in the characteristic polynomial of the  $j$ -th order Fibonacci sequence given by  $F_j(x) = x^j - x^{j-1} - \cdots - x - 1$ . It is known that the positive zeros of  $F_j(x)$  are of the form  $2 - O(2^{-j})$  [1],[2]. It has also been shown [8] that the single negative zero of  $F_j(x)$  has the form  $-1 + O(j^{-1})$  for  $j$  even and tends to  $-1$  monotonically as  $j \rightarrow \infty$ . By factoring these zeros one gets,

$$F_j(x) = (x - 2 + \varepsilon_j)(x + 1 - \delta_j)(x^{j-2} + a_{j-3}x^{j-3} + \cdots + a_1x + a_0),$$

where  $\delta_j$  and  $\varepsilon_j$  are positive, decreasing sequences for  $j = 4, 6, \dots$ . In [7] an explicit form for the coefficients  $a_i$  was found by solving a non-homogeneous linear recurrence relation of the form  $-a_n + b a_{n+1} + c a_{n+2} = 1$  where  $b = \varepsilon - 1 - \delta, c = (1 - \delta)(2 - \varepsilon)$ . As a byproduct of this solution, several combinatorial identities were formulated by solving by a standard method and comparing solutions; computer-generated and combinatorial proofs were also given to some identities, [4], [5], [6].

A recent paper, [13] has studied the asymptotics of the zeros of the positive and negative zeros of the derivatives and indefinite integrals of  $F_j$  as  $j \rightarrow \infty$  and shown this behavior is monotone for each derivative and integral. Moreover, one can write for

sufficiently large  $j$  and by factoring that

$$(1.1) \quad F_j^{(k)} = (x - 2 + \epsilon_j)(x + 1 - \delta_j) \sum_{i=0}^{j-(k+2)} a_i^{(k)} x^i.$$

Here  $F_j^{(k)}$  is the  $k^{th}$  derivative of  $F_j$ . Analogous to the case  $k = 0$  which corresponds to  $F_j$ ; in the present paper we derive recurrence relations of the form

$$(1.2) \quad -a_{k,n} - ba_{k,n+1} + ca_{k,n+2} = \binom{n+k+2}{n+2}, \quad k, n \geq 0,$$

with initial conditions. The proof of this result is by induction over  $k$  and is given in section 2. In sect. 3 we discuss the special case  $k = 1$  and  $c = b + 1$  which we call a singular solution and leads to an indeterminate form for  $b = -2$ . A result in [10] is shown to follow easily from an interesting identity. In sect. 4 we apply these ideas to finding closed forms for sums of powers of integers. More specifically, we show how one can find a closed form that contains sums of binomial coefficients with integer coefficients for the following sums,  $S_2(n) = \sum_{i=1}^n i^2$  and  $S_3(n) = \sum_{i=1}^n i^3$ .

## 2. Recurrence Relation and Identities

In the present section we extend some results in [3].

**Theorem 2.1.** *Define real numbers  $a_{k,n}, b, c$  such that  $k, n$  are nonnegative integers subject to initial conditions*

$$(2.1) \quad a_{0,0} = 1/c, \quad a_{0,1} = 1/c + b/c^2.$$

and  $a_{0,j}$  for  $j \geq 2$  by,

$$(2.2) \quad -a_{0,n} - ba_{0,n+1} + ca_{0,n+2} = 1, \quad n \geq 0,$$

and

$$(2.3) \quad a_{k+1,n} = \sum_{i=0}^n a_{k,i}.$$

Then we have

$$(2.4) \quad -a_{k,n} - ba_{k,n+1} + ca_{k,n+2} = \binom{n+k+2}{n+2}, \quad k, n \geq 0,$$

such that RHS of (2.4) comprise binomial coefficients in  $n, k$  relate the levels of the recurrence in  $k$ .



*Proof.* We prove inductively on  $k$ . Substituting (2.3) into LHS (2.4) with  $k = 1$  and employing initial conditions yields,

$$\begin{aligned}
 & - a_{1,n} - ba_{1,n+1} + ca_{1,n+2} \\
 = & - \sum_{i=0}^n a_{0,2i} - b \sum_{i=0}^{n+1} a_{0,i} + c \sum_{i=0}^{n+2} a_{0,i} \\
 = & - a_{0,0} - ba_{0,1} + ca_{0,2} - a_{0,1} - ba_{0,2} + ca_{0,3} - \dots \\
 & - a_{0,n} - ba_{0,n+1} + ca_{0,n+2} - ba_{0,0} + ca_{0,0} + ca_{0,1} \\
 = & n + 1 - b \cdot \frac{1}{c} + c \cdot \frac{1}{c} + c \cdot \left( \frac{1}{c} + \frac{b}{c^2} \right) = n + 3 = \binom{n+3}{n+2}.
 \end{aligned}$$

Thus, (2.4) is shown for  $k = 1$ . Next we note that,

**Lemma 2.1.**

$$(2.5) \quad a_{k,0} = \frac{1}{c},$$

$$(2.6) \quad a_{k,1} = \frac{k+1}{c} + \frac{b}{c^2}, \quad k \geq 0.$$

*Proof.* Both results follow from the initial conditions and by repeatedly applying (2.3), the former with  $n = 0$  and the latter with  $n = 1$ .  $\square$

We get, by applying the previous lemma and (2.3) to (2.4) with  $k = j + 1$ , the following

$$\begin{aligned}
 & - a_{j+1,n} - ba_{j+1,n+1} + ca_{j+1,n+2} \\
 = & - \sum_{i=0}^n a_{j,i} - b \sum_{i=0}^{n+1} a_{j,i} + c \sum_{i=0}^{n+2} a_{j,i} \\
 = & - a_{j,0} - ba_{j,1} + ca_{j,2} - a_{j,1} - ba_{j,2} + ca_{j,3} - \dots \\
 & - a_{j,n} - ba_{j,n+1} + ca_{j,n+2} - ba_{j,0} + ca_{j,0} + ca_{j,1} \\
 = & \binom{j+2}{2} + \binom{j+3}{3} + \dots + \binom{n+j+3}{n+3} - \frac{b}{c} + \frac{c}{c} + c \cdot \left( \frac{b}{c^2} + \frac{j+1}{c} \right) \\
 = & 1 + \binom{j+1}{1} + \binom{j+2}{2} + \binom{j+3}{3} + \dots + \binom{n+j+2}{n+2} \\
 = & \binom{n+j+3}{n+2},
 \end{aligned}$$

by a standard (ice or field hockey stick) result in binomial coefficients.  $\square$

Remark. The use of (2.3) to derive (2.4) is not obvious, but found by trial and error, nevertheless, it is somewhat close to the concept of superposition of families of solutions in linear differential equations.

It is also advantageous to write (2.4) in odd and even cases

**Corollary 2.1.**

$$(2.7) \quad -a_{k,2n} - ba_{k,2n+1} + ca_{k,2n+2} = \binom{2n+k+2}{2n+2},$$

$$(2.8) \quad -a_{k,2n+1} - ba_{k,2n+2} + ca_{k,2n+3} = \binom{2n+k+3}{2n+3}.$$

The following solutions [7] to (2.7,2.8) are given by

$$(2.9) \quad a_{0,2n+1} = \frac{1}{c} \left( \frac{1 - \left(\frac{1+b}{c}\right)^{n+1}}{1 - \frac{1+b}{c}} \right) + \frac{1}{c^{n+2}} \sum_{k=0}^n \sum_{i=2k+1}^{n+1+k} \binom{n+1+k}{i} \frac{b^i}{c^k},$$

$$(2.10) \quad a_{0,2n} = \frac{1}{c} \left( \frac{1 - \left(\frac{1+b}{c}\right)^{n+1}}{1 - \frac{1+b}{c}} \right) + \frac{1}{c^{n+2}} \sum_{k=0}^{n-1} \sum_{i=2k+2}^{n+1+k} \binom{n+1+k}{i} \frac{b^i}{c^k}.$$

Subtracting (2.9,2.10) gives

**Corollary 2.2.**

$$(2.11) \quad a_{0,2n+1} = a_{0,2n} + \frac{1}{c^{n+2}} \sum_{k=0}^n \binom{n+1+k}{2k+1} \frac{b^{2k+1}}{c^k}.$$

Moreover, let  $b = 1 - c$  in 2.11 and employing a result in [7] gives

**Corollary 2.3.**

$$(2.12) \quad \frac{1}{c^{n+2}} \sum_{k=0}^n \binom{n+1+k}{2k+1} \frac{(1-c)^{2k+1}}{c^k} = \frac{1/c^{2(n+1)} - 1}{1+c} = \frac{1}{c} \frac{1/c^{2(n+1)} - 1}{1+1/c}.$$

which LHS can be computed by Maple software and RHS is related to partial sum of geometric series. Setting  $c = -1$  in cor. 2.3 and employing L'Hôpital's rule produces a result in [10] given by

$$\sum_{k=0}^n (-1)^{n-k} 2^{2k} \binom{n+k+1}{2k+1} = n+1.$$

The derivation in [10] involved rational expressions and equating coefficients of infinite series. A few elegant results follow from 2.12. By setting  $c = 1/p$ ,  $p > 1$

$$\frac{p^2 - 1}{p^2} \sum_{k=0}^n \left( \frac{(p-1)^2}{p} \right)^k \binom{n+k+1}{2k+1} \approx p^n + O\left(\frac{1}{p^n}\right), n \geq 1.$$

and

$$(p^2 - 1)p^n \sum_{k=0}^n \left( \frac{(p-1)^2}{p} \right)^k \binom{n+k+1}{2k+1} + 1 = p^{2(n+1)},$$

and we see

**Corollary 2.4.** *If  $p$  is positive integer  $> 1$  and  $n \geq 1$  then*

$$\sum_{k=0}^n \left( \frac{(p-1)^2}{p} \right)^k \binom{n+k+1}{2k+1}$$

*is a positive, non-integer rational number.*

*Proof.* We have

$$\sum_{k=0}^n \left( \frac{(p-1)^2}{p} \right)^k \binom{n+k+1}{2k+1} = \frac{1}{p^n} (p^{2n} + p^{2(n-1)} + \dots + p^2 + 1),$$

and the result follows by considering the remainder *mod*  $p$ . □

One can also easily use Maple software to help compute the following: if  $p \neq 0, 1$

**Corollary 2.5.**

$$\sum_{k=0}^n \left( \frac{(p-1)^2}{p} \right)^k \binom{n+k+1}{2k} = \frac{1}{p^{n+1}} \left( \frac{p^{2n+3} + 1}{p + 1} - (p-1)^{2(n+1)} \right),$$

and also

**Corollary 2.6.**

$$\sum_{k=0}^n \left( \frac{(p-1)^2}{p} \right)^k \binom{n+k+2}{2k+1} = \frac{1}{p^{n+1}} \left( \frac{1 - p^{2n+4}}{1 - p^2} - (p-1)^{2(n+1)} \right).$$

It is noted that the possible elegance results from exponent 2 in the  $(p-1)^2$  term in the sums.

### 3. Singular solution: $k=1$

In the this section we consider (2.4) with  $k = 1$  so that,

$$(3.1) \quad -a_{1,n} - ba_{1,n+1} + ca_{1,n+2} = n + 3, \quad n \geq 0,$$

It is well-known that the standard solution for  $n$ , both odd and even, has the following (real) form,

$$(3.2) \quad a_{1,n} = A + Bn + Cn^2 + c_1\alpha^n + c_2\beta^n,$$

for some undetermined coefficients  $A, B, C$ , roots  $\alpha, \beta$  of the characteristic equation, and some constants  $c_1, c_2$ , that depend on the initial conditions. It is noted in (2.9, 2.10), the case  $c = b + 1$  leads to an indeterminate form. The case,  $c \neq b + 1$ , is considered in [7]. We have

$$(3.3) \quad -a_{1,n} - ba_{1,n+1} + (b+1) \cdot a_{1,n+2} = n + 3, \quad n \geq 0,$$

with initial conditions (2.1). The basic elementary idea is to exploit (2.4) with the solution (2.9, 2.10), and compare to the newly found solutions to (3.3) for odd and even cases. We solve (3.3) for solution  $a := a_{1,n}$  according to  $a = a_c + a_p$  such that,

$$\begin{aligned} -a_c - ba_c + (b+1) \cdot a_c &= 0, \\ -a_p - ba_p + (b+1)a_p &= n + 3. \end{aligned}$$

We have the following,

**Lemma 3.1.**

$$(3.4) \quad a := a_{1,n} = c_1 + c_2 \frac{(-1)^n}{(1+b)^n} + Bn + Cn^2,$$

where,

$$(3.5) \quad B = \frac{3b+8}{2(b+2)^2}, \quad C = \frac{1}{2(b+2)},$$

$$(3.6) \quad c_1 = \frac{b^2+5b+7}{(b+2)^3}, \quad c_2 = \frac{1}{(b+1) \cdot (b+2)^3}.$$

*Proof.* For  $a_c$ : The characteristic equation is given by:

$$(b+1) \cdot \alpha^2 - b \cdot \alpha - 1 = 0 \iff \alpha = 1, -\frac{1}{1+b}.$$

We next compute  $B, C$  by substituting  $a_p = Bn + Cn^2$  into (3.3) which yields the following equation,

$$\begin{aligned} & - C \cdot n^2 - B \cdot n - b \cdot (C \cdot (n+1)^2 + B \cdot (n+1)) \\ & - b \cdot (C \cdot (n+2)^2 + B \cdot (n+2)) = n+3, \end{aligned}$$

which, after simplification gives

$$(3.7) \quad n \cdot C \cdot (2b+4) + C \cdot (3b+4) + B \cdot (b+2) = n+3.$$

We obtain from (3.7), the pair of equations:

$$(3.8) \quad C \cdot (2b+4) = 1, \quad C \cdot (3b+4) + B \cdot (b+2) = 3.$$

Solving (3.8) for  $B, C$  yields (3.5). To solve for  $c_1, c_2$  in (3.4) use (3.5) with  $n = 0, 1$  and (2.6) with  $k = 1, c = b+1$ , to get the pair of equations

$$\begin{aligned} c_1 + c_2 &= \frac{1}{b+1}, \\ c_1 - \frac{c_2}{b+1} + \frac{1}{2(b+2)} + \frac{3b+8}{2(b+2)^2} &= \frac{3b+2}{(b+1)^2}, \end{aligned}$$

with solution (3.6). □

A result in [7],

**Corollary 3.1.**

$$\begin{aligned} (3.9) \quad a_{0,2n+1} &= \frac{b+3}{(b+2)^2} - \frac{1}{(1+b)^{2(n+1)}(b+2)^2} + \frac{2n+1}{b+2} \\ &= n+1 + \frac{1}{(1+b)^{n+2}} \sum_{k=0}^n \sum_{i=2k+1}^{n+1+k} \binom{n+1+k}{i} \frac{b^i}{(1+b)^k}, \end{aligned}$$

$$\begin{aligned} (3.10) \quad a_{0,2n} &= \frac{b+3}{(b+2)^2} + \frac{1}{(1+b)^{2(n+1)}(b+2)^2} + \frac{2n}{b+2} \\ &= n+1 + \frac{1}{(1+b)^{n+2}} \sum_{k=0}^{n-1} \sum_{i=2k+2}^{n+1+k} \binom{n+1+k}{i} \frac{b^i}{(1+b)^k}. \end{aligned}$$

We have analogous result for cor. 3.1. From 2.3 we have

$$(3.11) \quad a_{k+1,2n} = \sum_{i=0}^{2n} a_{k,i}, \quad a_{k+1,2n+1} = \sum_{i=0}^{2n+1} a_{k,i}.$$

We employ (3.11) in RHS of (3.9, 3.10) and use (3.4-3.6) to get after simplification,

**Corollary 3.2.**

$$(3.12) \quad \begin{aligned} a_{1,2n+1} &= \frac{b^2 + 5b + 7}{(b+2)^3} - \frac{1}{(1+b)^{2(n+1)}(b+2)^3} + \frac{(2n+1)^2}{2(b+2)} + \frac{(2n+1)(3b+8)}{2(b+2)^2} \\ &= \frac{(n+1)(n+2)}{b+1} + \sum_{i=1}^n \sum_{k=0}^{i-1} \sum_{j=2k+2}^{i+1+k} \binom{i+1+k}{j} \frac{b^j}{(1+b)^{i+2+k}} \\ &\quad + \sum_{i=0}^n \sum_{k=0}^i \sum_{j=2k+1}^{i+1+k} \binom{i+1+k}{j} \frac{b^j}{(1+b)^{i+2+k}}, \end{aligned}$$

$$(3.13) \quad \begin{aligned} a_{1,2n} &= \frac{b^2 + 5b + 7}{(b+2)^3} + \frac{1}{(1+b)^{2n+1}(b+2)^3} + \frac{2n^2}{b+2} + \frac{n(3b+8)}{(b+2)^2} \\ &= \frac{(n+1)^2}{b+1} + \sum_{i=1}^n \sum_{k=0}^{i-1} \sum_{j=2k+2}^{i+1+k} \binom{i+1+k}{j} \frac{b^j}{(1+b)^{i+2+k}} \\ &\quad + \sum_{i=0}^{n-1} \sum_{k=0}^i \sum_{j=2k+1}^{i+1+k} \binom{i+1+k}{j} \frac{b^j}{(1+b)^{i+2+k}}. \end{aligned}$$

One can then apply the similar procedure for  $k = 1, 2, \dots$  thereby producing the claimed sequence of identities. The combinatorial aspect of these identities is illustrated in the next section.

#### 4. $S_2(n), S_3(n)$ and binomial coefficients

The interesting case in (3.12, 3.13) is  $b \rightarrow -2$ . By continuity of RHS the limit exists and we note that binomial coefficients satisfy the following well-known equation

$$(4.1) \quad \binom{2n+2+k}{2n} - 2\binom{2n+3+k}{2n+1} + \binom{2n+4+k}{2n+2} = \binom{2n+2+k}{2n+2},$$

$$(4.2) \quad \binom{2n+3+k}{2n+1} - 2\binom{2n+4+k}{2n+2} + \binom{2n+5+k}{2n+3} = \binom{2n+3+k}{2n+3}.$$

By applying (4.1, 4.2) to (3.12, 3.13) we get that

**Corollary 4.1.**

$$\begin{aligned}
\lim_{b \rightarrow -2} & \frac{b^2 + 5b + 7}{(b+2)^3} - \frac{1}{(1+b)^{2(n+1)}(b+2)^3} + \frac{(2n+1)^2}{2(b+2)} + \frac{(2n+1)(3b+8)}{2(b+2)^2} \\
&= -\binom{2n+4}{2n+1}, \\
\lim_{b \rightarrow -2} & \frac{b^2 + 5b + 7}{(b+2)^3} + \frac{1}{(1+b)^{2n+1}(b+2)^3} + \frac{2n^2}{b+2} + \frac{n(3b+8)}{(b+2)^2} \\
&= -\binom{2n+3}{2n}.
\end{aligned}$$

Applying cor. 4.1 to (3.12, 3.13) we obtain

**Corollary 4.2.**

$$\begin{aligned}
(4.3) \quad \binom{2n+4}{2n+1} &= (n+1)(n+2) + \sum_{i=1}^n \sum_{k=0}^{i-1} \sum_{j=2k+2}^{i+1+k} \binom{i+1+k}{j} 2^j (-1)^{1+i+j+k} \\
&\quad + \sum_{i=0}^n \sum_{k=0}^i \sum_{j=2k+1}^{i+1+k} \binom{i+1+k}{j} 2^j (-1)^{1+i+j+k},
\end{aligned}$$

$$\begin{aligned}
(4.4) \quad \binom{2n+3}{2n} &= (n+1)^2 + \sum_{i=1}^n \sum_{k=0}^{i-1} \sum_{j=2k+2}^{i+1+k} \binom{i+1+k}{j} 2^j (-1)^{1+i+j+k} \\
&\quad + \sum_{i=0}^{n-1} \sum_{k=0}^i \sum_{j=2k+1}^{i+1+k} \binom{i+1+k}{j} 2^j (-1)^{1+i+j+k}.
\end{aligned}$$

We next show how it is possible to express sums of powers of integers of the form  $\sum i^p$ , where  $p = 2, 3$  as a sum of binomial coefficients. To see how this is done we have from [7]

$$(4.5) \quad n(n+1) = (-1)^{n+1} \sum_{k=0}^{n-1} \sum_{i=2k+2}^{n+1+k} \binom{n+1+k}{i} 2^{i-1} (-1)^{i+k}$$

$$(4.6) \quad (n+1)^2 = (-1)^{n+1} \sum_{k=0}^n \sum_{i=2k+1}^{n+1+k} \binom{n+1+k}{i} 2^{i-1} (-1)^{i+k}, \quad n \geq 0.$$

Employing (4.5, 4.6) in (4.3, 4.4) gives

$$(4.7) \quad \binom{2n+4}{2n+1} = (n+1)(n+2) + 2 \cdot \sum_{i=1}^n i \cdot (i+1) + 2 \cdot \sum_{i=0}^n (i+1)^2,$$

$$(4.8) \quad \binom{2n+3}{2n} = (n+1)^2 + 2 \cdot \sum_{i=1}^n i \cdot (i+1) + 2 \cdot \sum_{i=0}^{n-1} (i+1)^2.$$

Simplification of (4.7,4.8) leads to,

$$(4.9) \quad \binom{2n+4}{2n+1} = \sum_{i=0}^n (2i+2)^2,$$

$$(4.10) \quad \binom{2n+3}{2n} = \sum_{i=0}^n (2i+1)^2.$$

Summing (4.9, 4.10) yields,

$$\begin{aligned} \binom{2n+4}{2n+1} + \binom{2n+3}{2n} &= \sum_{i=1}^{2n+2} i^2, \\ \binom{2n+2}{2n-1} + \binom{2n+3}{2n} &= \sum_{i=1}^{2n+1} i^2. \end{aligned}$$

We also note the well-known formula

$$(4.11) \quad \sum_{i=1}^n i^2 = \frac{n}{6} \cdot (n+1) \cdot (2n+1) = \frac{1}{4} \binom{2n+2}{2n-1},$$

which is implied by (4.9).

Moreover, these ideas can be extended by applying (3.11,4.1,4.2) to corollary 3.2. We obtain, setting  $b = -2$  that

$$\begin{aligned} \binom{2n+4}{2n} &= -a_{2,2n} = - \left( \sum_{i=0}^n a_{1,2i} + \sum_{i=0}^{n-1} a_{1,2i+1} \right) \\ (4.12) \quad &= \sum_{i=0}^n (i+1)^2 - \sum_{i=1}^n \sum_{l=1}^i (-1)^l \cdot 2 \sum_{k=0}^{l-1} \sum_{j=2k+2}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1} \\ &\quad - \sum_{i=1}^n \sum_{l=0}^{i-1} (-1)^l \cdot 2 \sum_{k=0}^l \sum_{j=2k+1}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1} \\ &\quad + \sum_{i=0}^{n-1} (i+1)(i+2) - \sum_{i=1}^{n-1} \sum_{l=1}^i (-1)^l \cdot 2 \sum_{k=0}^{l-1} \sum_{j=2k+2}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1} \\ &\quad - \sum_{i=0}^{n-1} \sum_{l=0}^i (-1)^l \cdot 2 \sum_{k=0}^l \sum_{j=2k+1}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1}, \end{aligned}$$

## RECURRENCE RELATION WITH BINOMIAL COEFFICIENT

$$\begin{aligned}
(4.13) \quad \binom{2n+5}{2n+1} &= -a_{2,2n+1} = - \left( \sum_{i=0}^n a_{1,2i} + \sum_{i=0}^n a_{1,2i+1} \right) \\
&= \sum_{i=0}^n (i+1)^2 - \sum_{i=1}^n \sum_{l=1}^i (-1)^l \cdot 2 \sum_{k=0}^{l-1} \sum_{j=2k+2}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1} \\
&\quad - \sum_{i=1}^n \sum_{l=0}^{i-1} (-1)^l \cdot 2 \sum_{k=0}^l \sum_{j=2k+1}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1} \\
&\quad + \sum_{i=0}^n (i+1)(i+2) - \sum_{i=1}^n \sum_{l=1}^i (-1)^l \cdot 2 \sum_{k=0}^{l-1} \sum_{j=2k+2}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1} \\
&\quad - \sum_{i=0}^n \sum_{l=0}^i (-1)^l \cdot 2 \sum_{k=0}^l \sum_{j=2k+1}^{l+1+k} \binom{l+1+k}{j} (-1)^{j+k} 2^{j-1}.
\end{aligned}$$

We have by summation, from (4.5,4.6),

$$(4.14) \quad \sum_{l=1}^i l \cdot (l+1) = \sum_{l=1}^i (-1)^{l+1} \cdot \sum_{k=0}^{l-1} \sum_{j=2k+2}^{l+1+k} \binom{l+1+k}{j} 2^{j-1} (-1)^{j+k},$$

$$(4.15) \quad \sum_{l=1}^i (l+1)^2 = \sum_{l=1}^i (-1)^{l+1} \sum_{k=0}^l \sum_{j=2k+1}^{l+1+k} \binom{l+1+k}{j} 2^{j-1} (-1)^{j+k}.$$

Employing (4.14, 4.15) in (4.12, 4.13) we find that,

$$\begin{aligned}
(4.16) \quad \binom{2n+4}{2n} &= \sum_{i=1}^n \sum_{l=1}^i 2 \cdot l \cdot (l+1) + \sum_{i=1}^n \sum_{l=0}^{i-1} 2 \cdot (l+1)^2 + \sum_{i=0}^n (i+1)^2 \\
&\quad + \sum_{i=1}^{n-1} \sum_{l=1}^i 2 \cdot l \cdot (l+1) + \sum_{i=0}^{n-1} \sum_{l=0}^i 2 \cdot (l+1)^2 + \sum_{i=0}^{n-1} (i+1) \cdot (i+2),
\end{aligned}$$

$$\begin{aligned}
(4.17) \quad \binom{2n+5}{2n+1} &= \sum_{i=1}^n \sum_{l=1}^i 2 \cdot l \cdot (l+1) + \sum_{i=1}^n \sum_{l=0}^{i-1} 2 \cdot (l+1)^2 + \sum_{i=0}^n (i+1)^2 \\
&\quad + \sum_{i=1}^n \sum_{l=1}^i 2 \cdot l \cdot (l+1) + \sum_{i=0}^n \sum_{l=0}^i 2 \cdot (l+1)^2 + \sum_{i=0}^n (i+1) \cdot (i+2).
\end{aligned}$$



We consider (4.16,4.17) and find that

$$\begin{aligned}
 \binom{2n+4}{2n} &= 4 \cdot \sum_{i=1}^{n-1} \sum_{l=1}^i l \cdot (l+1) + 4 \cdot \sum_{i=1}^n \sum_{l=0}^{i-1} \cdot (l+1)^2 \\
 (4.18) \quad &+ \sum_{i=0}^n (i+1)^2 + 3 \sum_{i=0}^{n-1} (i+1)(i+2),
 \end{aligned}$$

$$\begin{aligned}
 \binom{2n+5}{2n+1} &= 4 \cdot \sum_{i=1}^n \sum_{l=1}^i l \cdot (l+1) + 4 \cdot \sum_{i=1}^n \sum_{l=0}^{i-1} \cdot (l+1)^2 \\
 (4.19) \quad &+ 3 \cdot \sum_{i=0}^n (i+1)^2 + \sum_{i=0}^n (i+1)(i+2).
 \end{aligned}$$

We simplify (4.18) employing (4.11) to get

$$\begin{aligned}
 \binom{2n+4}{2n} &= 8 \cdot \sum_{i=1}^{n-1} \sum_{l=1}^i l^2 + 4 \cdot \sum_{i=1}^{n-1} \sum_{l=1}^{i-1} l + \sum_{l=1}^n (8 \cdot l^2 + 3 \cdot l) + (n+1)^2 \\
 &= \sum_{i=1}^n \left[ 2 \cdot \binom{2i+2}{2i-1} + 4 \cdot \binom{i+1}{2} \right] + \binom{n+2}{2} \\
 (4.20) \quad &= \frac{1}{4} \cdot \sum_{i=0}^n \binom{4i+4}{3}.
 \end{aligned}$$

We find, after straightforward simplification of (4.20) that,

$$(4.21) \quad \sum_{i=1}^n (2i)^3 = 3 \cdot \binom{2n+4}{4} - \frac{9}{2} \cdot \binom{2n+2}{3} - 13 \cdot \binom{n+1}{2} - 3 \cdot \binom{n+1}{1}.$$

Using the fact that

$$\binom{2n+4}{2n} + \binom{2n+4}{2n+1} = \binom{2n+5}{2n+1}$$

and (4.9) we find that

$$\sum_{i=1}^n \left[ \frac{1}{3} (8 \cdot i^3 + 12 \cdot i^2 + 6 \cdot i + 1) + 6 \cdot i^2 + \frac{31}{3} \cdot i + \frac{14}{3} \right] + 5 = \binom{2n+5}{2n+1},$$

so that

$$\begin{aligned}
 \sum_{i=0}^n (2i+1)^3 &= 3 \cdot \binom{2n+5}{4} - \frac{9}{2} \cdot \binom{2n+2}{3} - 31 \cdot \binom{n+1}{2} - 14 \cdot \binom{n+1}{1}, n \geq 1. \\
 (4.22)
 \end{aligned}$$

Summing (4.21,4.22) gives

$$\begin{aligned} \sum_{i=1}^{2n+1} i^3 &= 3 \cdot \binom{2n+5}{4} + 3 \cdot \binom{2n+4}{4} - 9 \cdot \binom{2n+2}{3} - 44 \cdot \binom{n+1}{2} - 17 \cdot \binom{n+1}{1} \\ &= 3 \cdot \binom{2n+5}{4} + 3 \cdot \binom{2n+4}{4} - 9 \cdot \binom{2n+2}{3} - 44 \cdot \binom{n+1}{2} - 17 \cdot \binom{n}{1} - 17 \cdot \binom{n-1}{0}. \end{aligned} \quad (4.23)$$

We next use the fact that

$$(4.24) \quad (2n+1)^3 = 3 \left[ \binom{2n+2}{3} + \binom{2n+3}{3} \right] - 3 \binom{2(n+1)}{2} + \binom{2n+1}{1}.$$

Thus, for evenly many terms from (4.23, 4.24)

$$\begin{aligned} \sum_{i=1}^{2n} i^3 &= 3 \cdot \binom{2n+5}{4} + 3 \cdot \binom{2n+4}{4} - 9 \cdot \binom{2n+2}{3} - 44 \cdot \binom{n+1}{2} - 17 \cdot \binom{n+1}{1} \\ &\quad - 3 \left[ \binom{2n+2}{3} + \binom{2n+3}{3} \right] + 3 \binom{2(n+1)}{2} - \binom{2n+1}{1} \\ &= 3 \cdot \binom{2n+5}{4} + 3 \cdot \binom{2n+4}{4} - 3 \cdot \binom{2n+3}{3} - 12 \cdot \binom{2n+2}{3} + 3 \binom{2n+1}{2} \\ &\quad - 44 \cdot \binom{n+1}{2} - 13 \cdot \binom{n}{1} - 15 \cdot \binom{n-1}{0}. \end{aligned} \quad (4.25)$$

Finally, we present, without proof, a variation of a similar result in [9]:

$$\begin{aligned} S_p(n) - \sum_{j=1}^p c_j S_{p-j}(n) &= \int_0^n x^p dx, \\ c_j &= \frac{1}{p+1} \cdot \binom{p+1}{j+1} \cdot (-1)^{j+1}. \end{aligned}$$

## REFERENCES

- [1] Francois Dubeau, *On r-Generalized Fibonacci Numbers*, The Fibonacci Quarterly, **27:3**, (1989), 221-229.
- [2] Ivan Flores, *Direct Calculation of k-Generalized Fibonacci Numbers*, The Fibonacci Quarterly, **5:3**, (1967), 259-266.
- [3] George Grossman, *Linear recurrence relations and the binomial coefficients*, in Proceedings of XII<sup>th</sup> CZECH-POLISH-SLOVAK Mathematical School by the Faculty of Education of University J. E. Purkyně, Ústí nad Labem, Hubloš, June 2-4, 2005, pp. 111-119.
- [4] ———, Akalu Tefera and Aklulu Zeleke, *Summation Identities for Representation of Certain Real Numbers*, International Journal of Mathematics and Mathematical Sciences(e-journal), Volume 2006, Article ID 78739, 8 pages.

- [5] ———, Akalu Tefera and Aklulu Zeleke, *On proofs of certain combinatorial identities*, pre-print, 2004.
- [6] ———, Akalu Tefera and Aklulu Zeleke, *On Representation of Certain Real Numbers Using Combinatorial Identities*, pre-print, 2009.
- [7] ——— and Aklilu Zeleke, *On linear recurrence relations and combinatorial identities*, Journal of Concrete and Applicable Mathematics, Vol. 1, (2003), No. 3, pp. 229-245, Nova Science Publishers.
- [8] ——— and Sivaram Narayan, *On the characteristic polynomial of the  $j$ th order Fibonacci sequence*, Applications of Fibonacci numbers, Vol. 8, (1999), pp. 165-177 Kluwer Acad. Publ., Dordrecht.
- [9] Kenneth Ireland and Michael Rosen, *A Classical Introduction to Modern Number Theory*, Springer, 1990.
- [10] Georg Pólya and Gabor Szegő, *Aufgaben and Lehrsätze aus der Analysis I* New York, Dover Publications, 1945.
- [11] John Riordan. *Combinatorial Identities*, John Wiley & Sons Inc., 1968.
- [12] Marko Petkovšek, Herbert S. Wilf and Doron Zeilberger,  *$A = B$* , A. K. Peters, Wellesley, Massachusetts, 1996.
- [13] Xinyun Zhu and George Grossman, *On zeros of polynomial sequences*, JoCAA, (Journal of Computational Analysis with Applications,) **11**, No. 1, ( 2009), pp. 140-158.

DEPARTMENT OF MATHEMATICS, CENTRAL MICHIGAN UNIVERSITY, MT. PLEASANT, MI 48859  
GROSS1GW@CMICH.EDU

DEPARTMENT OF MATHEMATICS, LYMAN BRIGGS COLLEGE, EAST LANSING, MI 48825, ZELEKE@MSU.EDU

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF TEXAS OF THE PERMIAN BASIN, ODESSA, TX 79762, ZHU\_X@UTPB.EDU

# On the $q$ -extension of Genocchi polynomials

C. S. Ryoo

Department of Mathematics,  
Hannam University, Daejeon 306-791, Korea  
*e-mail: ryooos@hannam.ac.kr*

**Abstract :** In this paper we observe reflection symmetries of the  $q$ -extension of Genocchi polynomials,  $c_{n,q}(x)$ , using numerical investigation. By numerical experiments, we demonstrate a remarkably regular structure of the complex roots of the  $c_{n,q}(x)$  for  $q < 0$ . Finally, we give a table for the solutions of the  $q$ -extension of Genocchi polynomials.

**Key words :**  $q$ -extension of Genocchi numbers,  $q$ -extension of Genocchi polynomials, Roots of  $q$ -extension of Genocchi polynomials, Reflection symmetries of the  $q$ -extension of Genocchi polynomials

**2000 Mathematics Subject Classification :** 11S80, 11B68

## 1. Introduction

In [2], T. Kim constructed the  $q$ -extension of Genocchi numbers  $c_{n,q}$  and polynomials  $c_{n,q}(x)$  using generating functions. In order to study the  $q$ -extension of Genocchi polynomials  $c_{n,q}(x)$ , we must understand the structure of the  $q$ -extension of Genocchi polynomials  $c_{n,q}(x)$ . Therefore, using computer, a realistic study for the  $q$ -extension of Genocchi polynomials  $c_{n,q}(x)$  is very interesting. For related topics the interested reader is referred to [3, 4, 5, 6]. The main purpose of this paper is to consider reflection symmetries of the  $q$ -extension of Genocchi polynomials,  $c_{n,q}(x)$  for values of the index  $n$  by using computer. First, we introduce the  $q$ -extension of Genocchi polynomials  $c_{n,q}(x)$  (see [2, 3]). Let  $q$  be a complex number with  $|q| < 1$ . We consider the following generating functions:

$$F_q(t) = \sum_{n=0}^{\infty} c_{n,q} \frac{t^n}{n!} = e^{\frac{t}{1-q}} \sum_{n=0}^{\infty} \frac{(2n+1)(1-q^n)}{1-q^{2n+1}} \left( \frac{1}{1-q} \right)^{n-1} (-1)^{n-1} \frac{t^n}{n!}, \quad (1)$$

and

$$F_q(x, t) = \sum_{n=0}^{\infty} c_{n,q}(x) \frac{t^n}{n!} = e^{\frac{t}{1-q}} \sum_{n=0}^{\infty} \frac{(2n+1)(1-q^n)}{1-q^{2n+1}} \left( \frac{1}{1-q} \right)^{n-1} q^{nx} (-1)^{n-1} \frac{t^n}{n!}, \quad (2)$$

By simple calculation in (2), we have

$$c_{n,q}(x) = \sum_{i=0}^n \binom{n}{i} \frac{(2i+1)(1-q^i)}{1-q^{2i+1}} \left( \frac{1}{1-q} \right)^{n-1} (-1)^{i-1} q^{ix}.$$

When  $x = 0$ , we write  $c_{n,q} = c_{n,q}(0)$ , which are called the  $q$ -extension of Genocchi numbers.  $c_{n,q}(x)$  is a polynomial of degree  $= n$  in  $q^x$ . Since

$$\begin{aligned} \sum_{n=0}^{\infty} G_n(1-x) \frac{(-t)^n}{n!} &= F(1-x, -t) = \frac{-2t}{e^{-t} + 1} e^{(1-x)(-t)} \\ &= \frac{-2t}{e^t + 1} e^{xt} = -F(x, t) = -\sum_{n=0}^{\infty} G_n(x) \frac{t^n}{n!}, \end{aligned}$$

we obtain that

$$G_n(x) = (-1)^{n+1} G_n(1-x). \quad (3)$$

We prove that  $G_n(x), x \in \mathbb{C}$ , has  $Re(x) = \frac{1}{2}$  reflection symmetry in addition to the usual  $Im(x) = 0$  reflection symmetry analytic complex functions. The question is: what happens with the reflection symmetry (3), when one considers the  $q$ -extension of Genocchi polynomials? We are going now to reflection at  $\frac{1}{2}$  of  $x$  on the  $q$ -extension of Genocchi polynomials. For  $n \geq 0$ , we have

$$c_{n,q}^*(x) \equiv c_{n,q^{-1}}(1-x) = (-1)^{n-1} q^n c_{n,q}(x). \quad (4)$$

(4) is the  $q$ -analog of the classical reflection formula (3).  $c_{n,q}^*(x) (q > 0)$  has  $Im(x) = 0$  reflection symmetry analytic complex functions (Figure 3).  $c_{n,q}^*(x)$  has not  $Re(x) = 1/3$  reflection symmetry (Figure 3). If  $c_{n,q}^*(x) = 0 (q > 0)$ , then  $c_{n,q^{-1}}^*(1-x) = c_{n,q}^*(x^*) = c_{n,q^{-1}}^*(1-x^*) = 0$ , where  $*$  denotes complex conjugation.

## 2. Zeros of the $c_{n,q}^*(x)$

In order to study  $c_{n,q}^*(x)$ , we must understand the structure of the  $q$ -extension of Genocchi polynomials. In this section, by numerical investigation, we examine properties of the figures, look for patterns, and make open problems. First, we display the shapes of the  $c_{n,q}^*(x)$  and we investigate the zeros of the  $c_{n,q}^*(x)$ . For  $n = 1, \dots, 10$ , we can draw a plot of the  $c_{n,q}^*(x)$ , respectively. This shows the ten plots combined into one. We display the shape of  $c_{n,q}(x), c_{n,q}^*(x), n = 1, \dots, 10, -1 \leq x \leq 1$ .

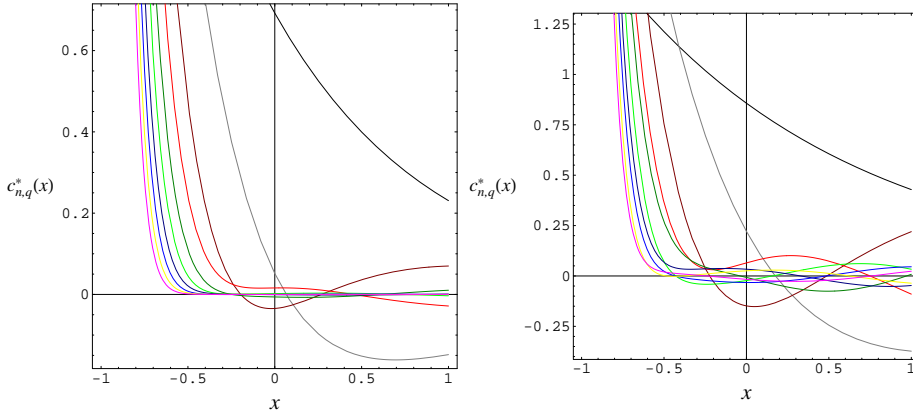


Figure 1: Curvers of  $c_{n,q}^*(x), q = 1/3$     Figure 2: Curvers of  $c_{n,q}^*(x), q = 1/2$

We investigate the beautiful zeros of the  $c_{n,q}^*(x)$  by using a computer. We plot the zeros of the  $c_{n,q}^*(x)$  for  $n = 30, 40, 50, 60, q = 1/3$  and  $x \in \mathbb{C}$ . (Figure 3). We plot the zeros of the  $c_{n,q}^*(x)$  for  $n = 30, 40, 50, 60, q = -1/3$  and  $x \in \mathbb{C}$ . (Figure 4). We plot the zeros of the  $c_{n,q}(x)$  for  $n = 30, 40, 50, 60, q = -1/3$  and  $x \in \mathbb{C}$ . (Figure 5).

We observe a remarkably regular structure of the complex roots of the  $c_{n,q}^*(x)$ . We hope to verify a remarkably regular structure of the complex roots of the  $c_{n,q}^*(x)$  (Table 1). Next, we calculate an approximate solution satisfying  $c_{n,q}^*(x), x \in \mathbb{R}$ . The results are given in Table 2.

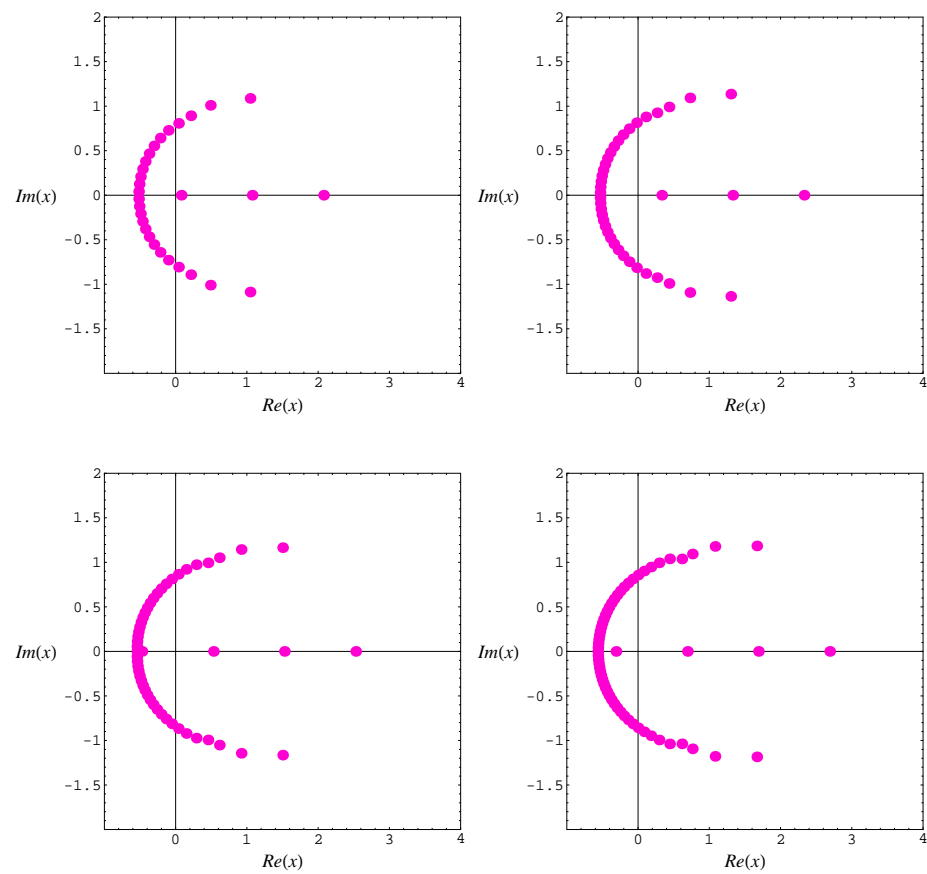
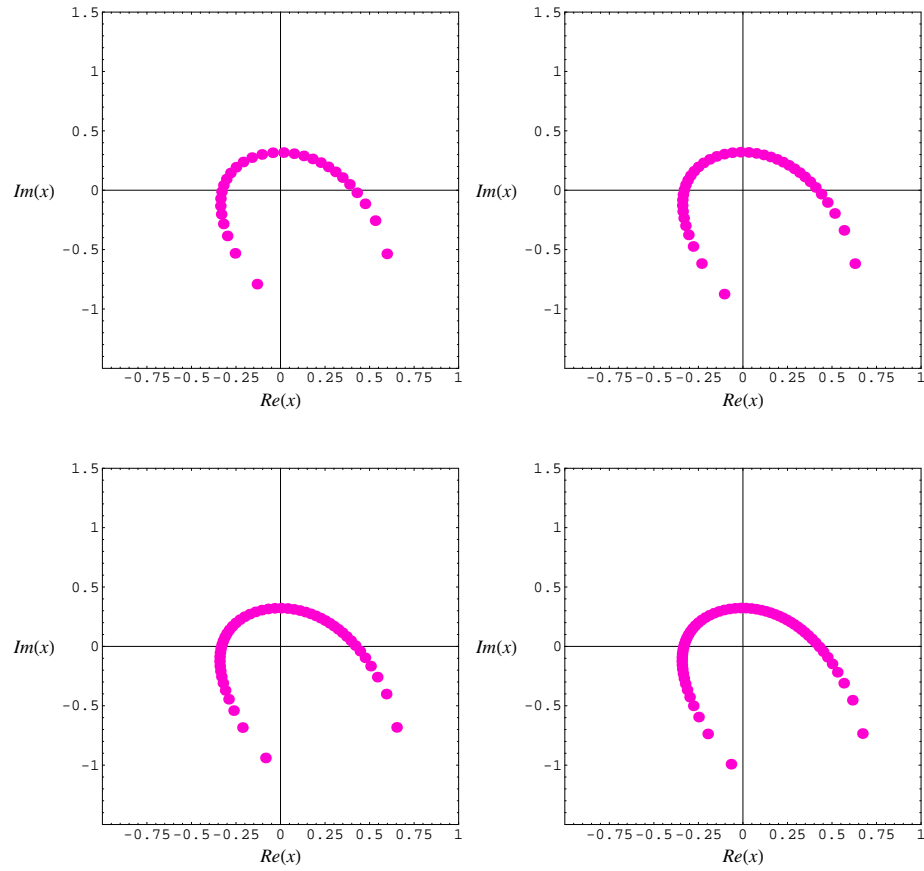


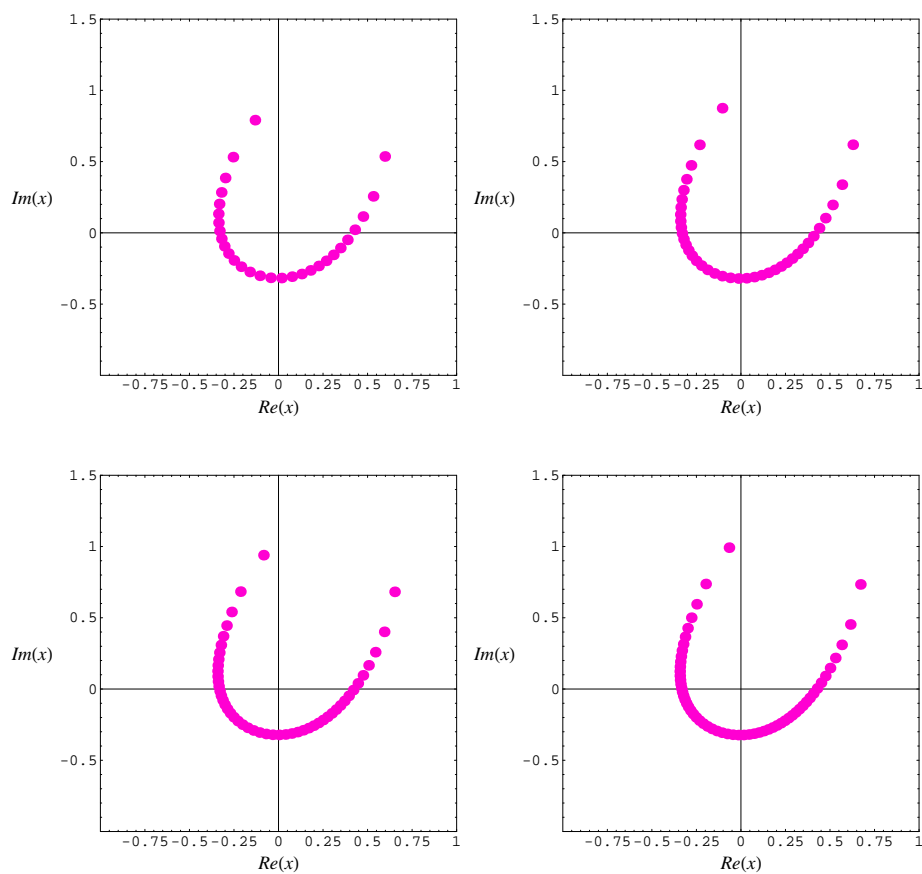
Figure 3: Zeros of  $c_{n,q}^*(x)$  for  $n = 10, 20, 30, 40, q = 1/3$

**Table 1.** Numbers of real and complex zeros of  $c_{n,q}^*(x)$

degree $n$	$q = -\frac{1}{3}$		$q = \frac{1}{3}$	
	real zeros	complex zeros	real zeros	complex zeros
2	0	1	1	0
4	0	3	1	2
6	0	5	1	4
8	0	7	3	4
10	0	9	3	6
12	0	11	3	8
14	0	13	3	10

Figure 4: Zeros of  $c_{n,q}^*(x)$  for  $n = 10, 20, 30, 40, q = -1/3$ **Table 2.** Approximate solutions of  $c_{n,q}^*(x) = 0, q = 1/3, x \in \mathbb{R}$ 

degree $n$	$x$
2	0.0653041
3	-0.19615, 0.268175
4	0.452437
5	-0.248071, 0.609023
6	0.743617
7	-0.144954, 0.861249
8	-0.35968, -0.0411608, 0.965587
9	0.0527129, 1.05928
10	-0.397998, 0.137934, 1.14427

Figure 5: Zeros of  $c_{n,q}(x)$  for  $n = 10, 20, 30, 40, q = -1/3$ **Table 3.** Approximate solutions of  $c_{n,q}(x) = 0, q = 1/3, x \in \mathbb{R}$ 

degree $n$	$x$
2	0.0653041
3	-0.19615, 0.268175
4	0.452437
5	-0.248071, 0.609023
6	0.743617
7	-0.144954, 0.861249
8	-0.35968, -0.0411608, 0.965587
9	0.0527129, 1.05928
10	-0.397998, 0.137934, 1.14427

We calculated an approximate solution satisfying  $c_{n,q}(x), c_{n,q}^*(x), q = -1/3, x \in \mathbb{R}$ . The results are



given in Table 4 and Table 5.

**Table 4.** Approximate solutions of  $c_{n,q}^*(x)$ ,  $q = -1/3$ ,  $x \in \mathbb{C}$

degree $n$	$x$
2	$-0.055099 + 0.157561i$
4	$-0.263443 - 0.136379i$ , $-0.082275 + 0.235274i$ , $0.291022 + 0.0575166i$
6	$-0.279677 + 0.0472548i$ , $-0.257121 - 0.289292i$ , $-0.0911906 + 0.260769i$ , $0.189278 + 0.211248i$ , $0.381371 - 0.0660123i$
8	$-0.309726 - 0.0831741i$ , $-0.260268 + 0.127805i$ , $-0.241352 - 0.386845i$ , $-0.095592 + 0.273355i$ , $0.123901 + 0.262149i$ , $0.294061 + 0.12797i$ , $0.429832 - 0.152132i$

**Table 5.** Approximate solutions of  $c_{n,q}(x)$ ,  $q = -1/3$ ,  $x \in \mathbb{C}$

degree $n$	$x$
2	$-0.055099 - 0.157561i$
4	$-0.263443 + 0.136379i$ , $-0.082275 - 0.235274i$ , $0.291022 - 0.0575166i$
6	$-0.279677 - 0.0472548i$ , $-0.257121 + 0.289292i$ , $-0.0911906 - 0.260769i$ , $0.189278 - 0.211248i$ , $0.381371 + 0.0660123i$
8	$-0.309726 + 0.0831741i$ , $-0.260268 - 0.127805i$ , $-0.241352 + 0.386845i$ , $-0.095592 - 0.273355i$ , $0.123901 - 0.262149i$ , $0.294061 - 0.12797i$ , $0.429832 + 0.152132i$

### 3. Directions for Further Research

Finally, we shall consider the more general problems. Prove or disprove:  $c_{n,q}^*(x) = 0$  has  $n - 1$  distinct solutions. Find the numbers of complex zeros  $C_{c_{n,q}^*(x)}$  of  $c_{n,q}^*(x)$ ,  $Im(x) \neq 0$ . Prove or disprove: Since  $n - 1$  is the degree of the polynomial  $c_{n,q}^*(x)$ , the number of real zeros  $R_{c_{n,q}^*(x)}$  lying on the real plane  $Im(x) = 0$  is then  $R_{c_{n,q}^*(x)} = n - 1 - C_{c_{n,q}^*(x)}$ , where  $C_{c_{n,q}^*(x)}$  denotes complex zeros. See Table 1 for tabulated values of  $R_{c_{n,q}^*(x)}$  and  $C_{c_{n,q}^*(x)}$ . The open question is: what happens with the reflection symmetry (4), when one considers the  $c_{n,q}^*(x)$  for  $q < 0$ ? (See Figures 4, 5). Find the equation of envelope curves bounding the real zeros lying on the plane. The author has no doubt that investigation along this line will lead to a new approach employing numerical method in the field of research of the  $c_{n,q}^*(x)$  to appear in mathematics and physics. The reader may refer to [3, 4, 5, 6] for the details

## References

- [1] T. KIM, On the  $q$ -extension of Euler and Genocchi numbers, *J. Math. Anal. Appl.*, **326** (2007), 1458-1465.
- [2] T. KIM, L.-C. JANG, H. K. PAK, A note on  $q$ -Euler and Genocchi numbers, *Proc. Japan Acad.*, **77 A** (2001), 139-141.
- [3] T. KIM, C. S. RYOO, Exploring the  $q$ -Euler numbers and polynomials, *Journal of concrete and Applicable Mathematics* , **7(4)** (2009), 349-357.
- [4] C.S.RYOO, A numerical computation on the structure of the roots of  $q$ -extension of Genocchi polynomials, *Applied Math. Letters*, **21** (2008), 348-354.
- [5] C.S.RYOO, Calculating zeros of the twisted Genocchi polynomials, *Advanced Studies in Contemporary Mathematics*, **17** (2008), 147-159.
- [6] C.S.RYOO, Y. S. YOO, A note on Euler numbers and polynomials, *Journal of concrete and Applicable Mathematics*, **7(4)** (2009), 341-348.

# On Best Simultaneous Approximation in Semi Metric Spaces

H. K. Pathak<sup>1</sup> and Satyaj Tiwari<sup>2</sup>

## Abstract

In this paper, the existence of invariant best simultaneous approximation in semi metric space is proved. In doing so, we have used a recent result of Moutawakil regarding the fixed points for set-valued mappings.

## 1. Introduction

In the realm of best approximation theory, it is vaible, meaningful and potentially productive to know whether some useful properties of the function being approximated is inherited by the approximating function. In this perspective, Meinardus [8] observed the general principle that could be applied, while doing so the author has employed a fixed point theorem as a tool to establish it. The result of Meinardus was further generalized by Habiniak [5], Smoluk [15] and Subrahmanyam [16].

On the other hand, Beg and Sahazad [2], Fan [4], Hicks and Humphries [6], Reich [10], Singh [13],[14] and many others have used fixed point theorems in approximation theory, to prove existence of best approximation. Various types of applications of fixed point theorems may be seen in Klee [7], Meinardus [8] and Vlasov [18]. Some applications of the fixed point theorems to best simultaneous approximation is given by Sahney and Singh [11]. For the detail survey of the subject we refer the reader to Cheney [3].

In this paper, we prove the existence of invariant best simultaneous approximation in semi metric space, while doing so, we use the recent result of

---

<sup>1</sup>Corresponding Author,E-mail:tsatyaj@yahoo.co.in

Moutawakil [19] on the fixed points for set-valued mappings.

## 2. Preliminaries and Definitions

Let  $(X, d)$  be a metric space. Let  $(CB(X), H)$  denote the hyperspace of nonempty closed bounded subsets of  $X$ , where  $H$  is the Housdroff metric induced by  $d$ , that is,

$$H(A, B) = \max\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\}$$

for all  $A, B \in CB(X)$ , where  $d(x, A) = \inf_{y \in A} \{d(x, y) : x \in X\}$  and  $A \subset X$ .

Although the fixed point theory for single valued maps is very rich and well developed, the multivalued case is not. Note that the multivalued mappings play a major role in many areas as in studying disjunctive logic programs.

On the other hand; it has been observed that (see example [22]) that the distance function used in certain metric theorems proofs need not satisfy the triangular inequality nor  $d(x, x) = 0$  for all  $x$ . Motivated by this fact, Hicks and Rhoades [22] established some common fixed point theorems in symmetric spaces and proved that very general probabilistic structures admits a compatible symmetric or semi-metric. Recall that a symmetric on a set  $X$  is a non negative real valued function  $d$  on  $X \times X$  such that (i)  $d(x, y) = 0$  if and only if  $x = y$ , (ii)  $d(x, y) = d(y, x)$ .

In order to unify the notation, we need the following notation:

(W.4) Given  $\{x_n\}, \{y_n\}$  and  $x$  in  $X$ ,  $\lim_{n \rightarrow \infty} d(x_n, x) = 0$  and  $\lim_{n \rightarrow \infty} d(x_n, y_n) = 0$  imply that  $\lim_{n \rightarrow \infty} d(y_n, x) = 0$ .

A sequence in  $X$  is called a  $d$ -cauchy sequence if it satisfies the usual metric condition.  $X$  is  $S$ -complete if for every  $d$ -cauchy sequence  $(x_n)$ , there exists  $x$  in  $X$  with  $\lim_{n \rightarrow \infty} d(x_n, x) = 0$

### The Hausdorff distance in a symmetric space.

**Definition.** Let  $(X, d)$  be a symmetric space and  $A$  a nonempty subset of  $X$ .

(i) We say that  $A$  is  $d$ -closed iff  $\bar{A}^d = A$  where

$$\bar{A}^d = \{x \in X : D(x, A) = 0\} \text{ and } d(x, A) = \inf\{d(x, y) : y \in A\}$$

(ii) We say that  $A$  is  $d$ -bounded iff  $\delta_d(A) < \infty$  where  $\delta_d(A) = \sup\{d(x, y) : x, y \in A\}$

The following definition is a generalization of the well-known Hausdorff distance to the setting of symmetric case.

**Definition.** Let  $(X, d)$  be a  $d$ -bounded symmetric space and let  $C(X)$  be the set of all nonempty  $d$ -closed subset of  $(X, d)$ . Consider the function  $H : 2^X \times 2^X \rightarrow R$  defined by

$$H(A, B) = \max\left\{\sup_{a \in A} d(a, B), \sup_{b \in B} d(b, A)\right\}$$

for all  $A, B \in C(X)$ .

**Remark.** It is easy to see that  $(C(X), D)$  is a symmetric space.

Now we give the notion of convex structure introduced by Gudder [20](see also, Petrusel [21]).

Let  $X$  be a set and  $F : [0, 1] \times X \times X \rightarrow X$  a mapping. Then the pair  $(X, F)$  forms a *convex prestructure*. Let  $(X, F)$  be a convex prestructure. If  $F$  satisfies the following conditions:

- (i)  $F(\lambda, x, F(\mu, y, z)) = F(\lambda + (1 - \lambda)\mu, F(\lambda(\lambda + (1 - \lambda)\mu)^{-1}, x, y), z)$  for every  $\lambda, \mu \in (0, 1)$  with  $\lambda + (1 - \lambda)\mu \neq 0$  and  $x, y, z \in X$ .
- (ii)  $F(\lambda, x, x) = x$  for any  $x \in X$  and  $\lambda \in (0, 1)$ ,

then  $(X, F)$  forms a *semi-convex structure*. If  $(X, F)$  is a semi-convex structure, then

$$(SC1) \quad F(1, x, y) = x \text{ for any } x, y \in X.$$

A semi-convex structure is said to be *regular* if

$$(SC2) \quad \lambda \leq \mu \Rightarrow F(\lambda, x, y) \leq F(\mu, x, y) \text{ where } \lambda, \mu \in (0, 1).$$

A semi-convex structure  $(X, F)$  is said to form a *convex structure* if  $F$  also satisfies the conditions

- (iii)  $F(\lambda, x, y) = F(1 - \lambda, y, x)$  for every  $\lambda \in (0, 1)$  and  $x, y \in X$ .
- (iv) if  $F(\lambda, x, y) = F(\lambda, x, z)$  for some  $\lambda \neq 1, x \in X$  then  $y = z$ .

Let  $(X, F)$  be a convex structure. A subset  $Y$  of  $X$  is called (a) *F-starshaped* if there exist  $p \in Y$  so that for any  $x \in Y$  and  $\lambda \in (0, 1)$ ,  $F(\lambda, x, p) \in Y$ . (b) *F-convex* if for any  $x, y$  in  $Y$  and  $\lambda \in (0, 1)$ ,  $F(\lambda, x, y) \in Y$ . For  $F(\lambda, x, y) = \lambda x + (1 - \lambda)y$ , we obtain the known notion of starshaped convexity from linear spaces. Petrusel [20] noted with an example that a set can be a *F*-semi convex structure without being a convex structure. Let  $(X, F)$  be a semi-convex structure. A subset  $Y$  of  $X$  is called *F semi-starshaped* if there exists  $p \in Y$  so that for any  $x \in Y$  and  $\lambda \in (0, 1)$ ,  $F(\lambda, x, p) \in Y$ . A Banach space  $X$  with semi-convex structure  $F$  is said to satisfy condition  $(P_1)$  at  $p \in K$  (where  $K$  is semi-starshaped and  $p$  is star centre) if  $F$  is continuous relative to the following argument : for any  $x, y \in X, \lambda \in (0, 1)$

$$\| (F(\lambda, x, p) - F(\lambda, y, p)) \| \leq \lambda \| x - y \| .$$

To prove our main result(Theorem 3.1 below), we shall make use of the following result due to Moutawakil ([19], theorem 2.2.1).

**Theorem A.** Let  $(X, d)$  be a  $d$ -bounded and  $S$ -complete symmetric space satisfying  $(W-4)$  and let  $T : X \rightarrow C(X)$  be a set-valued mapping such that

$$H(Tx, Ty) \leq kd(x, y)$$

for all  $x, y \in X$ , where  $k \in [0, 1)$ . Then there exist  $u \in X$  such that  $u \in Tu$ .

Let  $(X, d)$  be a metric space and  $G$  a nonempty subset of  $X$ . Suppose  $A \in C(X)$ , the set of nonempty  $d$ -closed subsets of  $(X, d)$ , then we write

$$r_G(A) = \inf_{g \in G} \sup_{a \in A} d(a, g)$$

$$cent_G(A) = \{g_0 \in G : \sup_{a \in A} d(a, g_0) = r_G(A)\}.$$

The number  $r_G(A)$  is called the Chebyshev radius of  $A$  w.r.t  $G$  and an element  $y_0 \in cent_G(A)$  is called a best simultaneous approximation of  $A$  w.r.t  $G$ . If  $A = \{x\}$ , then  $r_G(A) = d(x, G)$  and  $cent_G(A)$  is the set of all best approximations of  $x$  out of  $G$ . We also refer the reader to Milman [9] for further details.

### 3. Main Results

Now we state and prove our main result.

**Theorem 3.1** Let  $(X, d)$  be a  $d$ -bounded and  $S$ -complete semi-metric space with semi-convex structure satisfying condition  $(P_1)$  and  $T : X \rightarrow C(X)$  be a set-valued mapping. Let  $G \in C(X)$ . For  $A \subset C(X)$ , if  $cent_G(A)$  is nonempty, compact, semi-starshaped,  $T$ -invariant and  $T$  satisfy the following conditions:

(i)  $T$  is continuous on  $cent_G(A)$ , and

(ii)  $H(Tx, Ty) \leq kd(x, y)$

for all  $x, y \in cent_G(A)$  with  $x \neq y, d(x, y) > 0$ , then  $cent_G(A)$  contains a  $T$ -invariant point.

**Proof.** Let  $p$  be the starcentre of  $cent_G(A)$ . Then  $F(x, p, \lambda) \in cent_G(A)$  for each  $x \in cent_G(A)$ . Let  $\{k_n\}_{n=1}^\infty$  be a real sequence with  $0 \leq k_n < 1$  such that  $k_n \rightarrow 1$  as  $n \rightarrow \infty$ . Define  $T_n : cent_G(A) \rightarrow C(cent_G(A))$  by

$$T_n x = F(k_n, Tx, p) = \bigcup_{y \in Tx} F(k_n, y, p)$$

for all  $x \in cent_G(A)$ . Since  $p$  is semi-starcenter of  $cent_G(A)$  and  $T(cent_G(A)) \subseteq cent_G(A)$ , It follows that  $T_n$  maps  $cent_G(A)$  to itself for each  $n$ . Now applying condition  $(P_1)$ , we obtain

$$\begin{aligned} H(T_n x, T_n y) &= H(F(k_n, Tx, p), F(k_n, Ty, p)) \\ &\leq k_n H(Tx, Ty) \\ &\leq k'_n d(x, y) \end{aligned}$$

i.e.,

$$H(T_n x, T_n y) \leq k'_n d(x, y)$$

for all  $x, y \in cent_G(A)$ . It follows by Theorem A that each  $T_n$  has a fixed point, say  $z_n$ . Since  $cent_G(A)$  is complete,  $\{z_n\}$  has a convergent subsequence  $\{z_{n_i}\}$  such that  $z_{n_i} \rightarrow z$  (say) as  $i \rightarrow \infty$  for some  $z \in X$ . Since

$$z_{n_i} \in T_{n_i} z_{n_i} = F(k_{n_i}, T z_{n_i}, p)$$

and  $k_{n_i} \rightarrow 1$  as  $i \rightarrow \infty$ , it follows that  $z \in Tz$ . Hence  $\text{cent}_G(A)$  contains a  $T$ -invariant point. This completes the proof.

**Remark 3.2** Let  $(X, d)$  be a  $d$ -bounded and  $S$ -complete semi-metric space with semi-convex structure satisfying condition  $(P_1)$  and let  $T$  be a self map on  $X$ . Let  $G \in X$ . For  $A \subset X$ , if  $\text{cent}_G(A)$  is nonempty, compact, semi-starshaped,  $T$ -invariant and  $T$  satisfy the following conditions:

(i)  $T$  is continuous on  $\text{cent}_G(A)$ , and

(ii)  $(Tx, Ty) \leq kd(x, y)$

for all  $x, y \in \text{cent}_G(A)$  with  $x \neq y, d(x, y) > 0$ , then  $\text{cent}_G(A)$  contains a  $T$ -invariant point.

## References

- [1] Beg, I. and Azam, A., Fixed points of multivalued locally contractive mappings, Boll. U.M.I., (7) 4-A (1990), 227-233.
- [2] Beg, I. and Shahzad, N., An application of a fixed point theorem to best approximation, Approx. Theory and its Appl., 10:3(1994), 1-4.
- [3] Cheney, E. W., Application of fixed point theorems to approximation theory, Theory of Approximations, Academic Press (1976), 1-8.
- [4] Fan, Ky., Extension of two fixed point theorems of F.E.Browder, Math Z., 112 (1969), 234-240.
- [5] Habiniak, L., Fixed point theorems and invariant approximations, J. Approximation Theory, 56(1989), 241-244.
- [6] Hicks, T.L. and Humphries M.D., A note on fixed point theorems, J. Approximation Theory, 34 (1982), 221-222.
- [7] Klee, V., Convexity of chebyshev sets, Math. Ann., 142(1961), 292-304.
- [8] Meinardus, G., Invariant bei Lineare Approximation, Arch. Rational Mech. Anal., 14(1963), 301-303.
- [9] Milman, P.D., On best simultaneous approximation in normed linear spaces, J. Approximation Theory, 20(1977), 223-238.



- [10] Reich, S., Approximate selection, best approximations, fixed points and invariant sets, *J. Math. Anal. Appl.*, 62(1978), 104-113.
- [11] Sahney, B.N. and Singh S.P., On best simultaneous approximation, *Approximation Theory III*, Academic Press (1980), 783-789.
- [12] Singh, S.P., Application of fixed point theorems in approximation theory, *Applied Nonlinear Analysis*, Academic Press (1979), 389-394.
- [13] ———, Application of a fixed point theorem to approximation theory, *J. Approx. Theory*, 25(1979), 88-89.
- [14] ———, Some results on best approximation in locally convex spaces, *J. Approx. Theory*, 28(1980), 72-76.
- [15] Smoluk, A., Invariant approximations, *Mathematyka [Polish]*, 17(1981), 17-22.
- [16] Subrahmanyam, P.V., An application of a fixed point theorem to best approximations, *J. Approx. Theory*, 20(1977), 165-172.
- [17] Takahashi, W., A convexity in metric spaces and nonexpansive mappings I, *Kodai Math. sem. Rep.*, 22(1970), 142-149.
- [18] Vlasov, L.P., Chebyshev sets in Banach spaces, *Soviet Math. Polody*, 2(1961), 1373-1374.
- [19] Driss El Moutawakil, A fixed point Theorem for multivalued maps in symmetric spaces, *Applied Mathematics E-Notes*, 4(2004), 26-32.
- [20] Gudder, S.P., A general theory of convexity, *Rend. Sem. Mat. Milano*, 49,(1979), 89-96.
- [21] Petrusel, A., Starshaped and fixed points, *Seminar on fixed point theory (Cluj-Napoca)*, *Stud. Univ. "Babes-Bolyai"*, Nr.3,(1987), 19-24.
- [22] T.L. Hicks and B.E. Rhoades, Fixed point theory in symmetric spaces with applications to probabilistic spaces, *Nonlinear Analysis* 36 (1999), 331-344.

<sup>1</sup>H. K. Pathak  
 School of Studies in Mathematics  
 Pt. Ravishankar Shukla University, Raipur  
 (C.G) 492010, India

E-mail : [hkpathak@sify.com](mailto:hkpathak@sify.com)

<sup>2</sup>Satyaj Tiwari

Department of Mathematics

Shri Shankaracharya Institute of Professional Management and Technology,

P.O. Sejabahar, Raipur

(C.G) 490021, India

E-mail : [tsatyaj@yahoo.co.in](mailto:tsatyaj@yahoo.co.in)

# ON BEST UNIFORM APPROXIMATION OF PERIODIC FUNCTIONS BY TRIGONOMETRIC POLYNOMIALS

MICHAEL I. GANZBURG

**ABSTRACT.** We discuss inequalities of the form  $C_1 E_n(f')_{L_{2\pi}^*} \leq E_n(f)_{C_{2\pi}^*} \leq C_2 E_n(f')_{L_{2\pi}^*}$  between the errors of approximation in the uniform and integral metrics of periodic functions  $f$  and  $f'$  by trigonometric polynomials. As applications, we obtain upper and lower estimates for  $E_n(f)_{C_{2\pi}^*}$  in terms of the Fourier coefficients of a function  $f$ .

## 1. INTRODUCTION

Let  $C_{2\pi}^*$  be the Banach space of all  $2\pi$ -periodic continuous functions  $f$  on the real line  $\mathbb{R}$  with the finite norm  $\|f\|_{C_{2\pi}^*} := \max_{x \in [0, 2\pi)} |f(x)|$ ;  $L_{2\pi}^*$  the Banach space of all  $2\pi$ -periodic measurable functions  $f$  on  $[0, 2\pi)$  with the finite norm  $\|f\|_{L_{2\pi}^*} := \int_0^{2\pi} |f(x)| dx$ , and  $\mathcal{T}_n$  the class of all trigonometric polynomials of degree  $\leq n$ . For  $n = 0, 1, \dots$ , we define the approximation errors by

$$E_n(f)_{C_{2\pi}^*} := \inf_{T_n \in \mathcal{T}_n} \|f - T_n\|_{C_{2\pi}^*}, \quad E_n(f)_{L_{2\pi}^*} := \inf_{T_n \in \mathcal{T}_n} \|f - T_n\|_{L_{2\pi}^*}.$$

Throughout the paper  $C, C_1, \dots$  denote positive constants independent of  $n$ . The same symbol does not necessarily denote the same constant in different occurrences.

In this paper we discuss upper and lower estimates for the error of uniform approximation of periodic functions by trigonometric polynomials in terms of their Fourier coefficients. The estimates for  $E_n(f)_{C_{2\pi}^*}$  are based on inequalities of the form

$$C_1(n) E_n(f')_{L_{2\pi}^*} \leq E_n(f)_{C_{2\pi}^*} \leq C_2(n) E_n(f')_{L_{2\pi}^*},$$

where  $C_2(n) = 1/2$ ,  $n \geq 1$  (Theorem 2.1) and  $C_1(n) = [4(n+1)]^{-1}$ ,  $n \geq 1$  (Theorem 3.1). Our approach is illustrated by five example.

## 2. UPPER ESTIMATES

We first discuss upper estimates for  $E_n(f)_{C_{2\pi}^*}$  in terms of the Fourier coefficients of an individual function  $f$ . Jackson's theorems are inapplicable in this case, while various linear approximation methods have proved to be efficient [8], [1] [9], [4]. In particular, if the series

$$f(x) = a_0/2 + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) \quad (2.1)$$

---

*Key words and phrases.* Best approximation, trigonometric polynomials, Fourier coefficients.

is absolutely convergent, then the following trivial estimate

$$E_n(f)_{C_{2\pi}^*} \leq \sum_{k=n+1}^{\infty} (|a_k| + |b_k|)$$

generated by the Fourier approximation can be efficient for rapidly decreasing Fourier coefficients [4] as well as for slowly decreasing ones. The latter statement is supported by the estimate  $E_n(f_\lambda)_{C_{2\pi}^*} \leq C(n+1)^{-\lambda+1}$ , where  $f_\lambda(x) := \sum_{k=1}^{\infty} k^{-\lambda} \cos kx$ ,  $\lambda > 1$ . Lower estimate (3.14) given in Section 3 shows that the upper estimate cannot be improved. Note that for an integer  $\lambda > 1$  an asymptotic formula for  $E_n(f_\lambda)_{C_{2\pi}^*}$  was found in [2].

Here we consider upper estimates for  $E_n(f)_{C_{2\pi}^*}$ , which can be applied to conditionally and absolutely convergent series (2.1). Our approach is based on the following simple result for functions of bounded variation.

**Theorem 2.1.** *Let  $f \in C_{2\pi}^*$  be an absolutely continuous function with  $f' \in L_{2\pi}^*$ . Then the following inequality holds:*

$$E_n(f)_{C_{2\pi}^*} \leq (1/2)E_n(f')_{L_{2\pi}^*}. \quad (2.2)$$

*Proof.* Let  $B_1(x) := \sum_{k=1}^{\infty} k^{-1} \sin kx$  be a Bernoulli function and let  $Q_n \in \mathcal{T}_n$  be a polynomial of best mean approximation to  $f'$ . Then the following relation holds [5, Eq. 1.5.1]:

$$f(x) = a_0/2 + \frac{1}{\pi} \int_0^{2\pi} B_1(x-t)f'(t)dt.$$

In addition, the function

$$T_n(x) := a_0/2 + \frac{1}{\pi} \int_0^{2\pi} B_1(x-t)Q_n(t)dt$$

belongs to  $\mathcal{T}_n$ . Therefore,

$$\begin{aligned} E_n(f)_{C_{2\pi}^*} \leq \|f - T_n\|_{C_{2\pi}^*} &\leq \frac{1}{\pi} \sup_{t \in [0, 2\pi)} |B_1(t)| \int_0^{2\pi} |f'(t) - Q_n(t)| dt \\ &= (1/2)E_n(f')_{L_{2\pi}^*}. \end{aligned}$$

This proves (2.2).  $\square$

There are several results on finding or estimating  $E_n(g)_{L_{2\pi}^*}$ . Three of them are given in the following lemma.

**Lemma 2.1.** (a) *If a sequence  $\{B_k\}_{k=n+1}^{\infty}$  of the Fourier coefficients of an odd function  $g(x) = \sum_{k=1}^{\infty} B_k \sin kx \in L_{2\pi}^*$  is 2-monotone, that is,  $B_k > 0$ ,  $B_k - B_{k+1} \leq 0$ , and  $B_k - 2B_{k+1} + B_{k+2} \geq 0$  for  $k \geq n+1$ , then  $g \in L_{2\pi}^*$  and*

$$E_n(g)_{L_{2\pi}^*} = 4 \sum_{m=0}^{\infty} \frac{B_{(2m+1)(n+1)}}{2m+1}.$$

(b) *If a sequence  $\{A_k\}_{k=n+1}^{\infty}$  of the Fourier coefficients of an even function  $g(x) = A_0/2 + \sum_{k=1}^{\infty} A_k \cos kx$  is 3-monotone, that is,  $A_k > 0$ ,  $A_k - A_{k+1} \leq 0$ ,  $A_k -$*

$2A_{k+1} + A_{k+2} \geq 0$  and  $A_k - 3A_{k+1} + 3A_{k+2} - A_{k+3} \leq 0$  for  $k \geq n+1$ , then  $g \in L_{2\pi}^*$  and

$$E_n(g)_{L_{2\pi}^*} = 4 \sum_{m=0}^{\infty} (-1)^m \frac{A_{(2m+1)(n+1)}}{2m+1}.$$

(c) If  $g(x) = A_0/2 + \sum_{k=1}^{\infty} A_k \cos kx \in L_{2\pi}^*$ , then

$$E_n(g)_{L_{2\pi}^*} \leq C \left( |A_{n+1}| + |A_{2n+2}| + \sum_{k=1}^{\infty} k |A_{k+n} - 2A_{k+n+1} + A_{k+n+2}| \right).$$

Statements (a) and (b) of Lemma 2.1 were proved by Nagy [6] (see also [8, Sections 2.11.5 and 2.13.32]), while statement (c) was established by the author [4, Lemma 3.4]. Note that the condition  $g \in L_{2\pi}^*$  in Lemma 2.1(a) is equivalent to the statement that  $g$  is the Fourier series and, in addition, it is equivalent to convergence of the series  $\sum_{k=1}^{\infty} B_k/k$ .

Combining Theorem 2.1 and Lemma 2.1, we obtain upper estimates for  $E_n(f)_{C_{2\pi}^*}$ . Let us illustrate this approach by the following examples.

**Example 2.1.**  $f(x) = \sum_{k=2}^{\infty} \frac{\sin kx}{k \log^q k}$ ,  $q > 0$ . Then the series  $f'(x) = \sum_{k=2}^{\infty} \frac{\cos kx}{\log^q k}$  converges at every  $x \in (0, 2\pi)$  and  $f' \in L_{2\pi}^*$  since  $\{\log^{-q} k\}_{k=2}^{\infty}$  is a convex sequence [10, Theorem 5.1.5]. Moreover,  $\{\log^{-q} k\}_{k=2}^{\infty}$  is a 3-monotone sequence. Therefore by Theorem 2.1 and Lemma 2.1(b), we obtain for  $n \geq 1$ ,

$$\begin{aligned} E_n(f)_{C_{2\pi}^*} &\leq (1/2) E_n(f')_{L_{2\pi}^*} \leq 2 \sum_{m=0}^{\infty} \frac{(-1)^m}{(2m+1) \log^q[(2m+1)(n+1)]} \\ &\leq 2 \log^{-q}(n+1). \end{aligned} \quad (2.3)$$

Note that in case  $q \in (0, 1]$  the Fourier series for  $f$  is conditionally convergent.

In addition, we remark that an estimate

$$E_n(f)_{C_{2\pi}^*} \leq C \log^{-q}(n+1) \quad (2.4)$$

can be deduced from Lemma 1(c) as well. Indeed, setting  $A_k := \log^{-q} k$ ,  $k \geq 2$ , we have for  $n \geq 2$

$$\begin{aligned} \sum_{k=n+1}^{\infty} k |A_{k+n} - 2A_{k+n+1} + A_{k+n+2}| &= \sum_{k=n+1}^{\infty} (k-n) |A_k - 2A_{k+1} + A_{k+2}| \\ &\leq \sum_{k=n+1}^{\infty} k \max_{k \leq y \leq k+2} |d^2(\log^{-q} y)/dy^2| \leq C \sum_{k=n+1}^{\infty} k^{-1} \log^{-(q+1)} k \\ &\leq C \log^{-q}(n+1). \end{aligned}$$

This implies (2.4).

**Example 2.2.**  $f(x) = \sum_{k=1}^{\infty} k^{\lambda} \exp(-A k^{\alpha}) \cos kx$ ,  $\lambda \in \mathbb{R}$ ,  $A > 0$ ,  $\alpha > 0$ . Then  $f'(x) = -\sum_{k=1}^{\infty} k^{\lambda+1} \exp(-A k^{\alpha}) \sin kx \in L_{2\pi}^*$ . Since for  $n_0$  large enough, the

sequence  $\{k^{\lambda+1} \exp(-A k^\alpha)\}_{k=n+1}^\infty$  is 2-monotone for  $n \geq n_0$ , we obtain from Theorem 2.1 and Lemma 2.1(a) for  $n \geq n_0$

$$\begin{aligned}
E_n(f)_{C_{2\pi}^*} &\leq (1/2)E_n(f')_{L_{2\pi}^*} = 2(n+1)^{\lambda+1} \\
&\times \left( \exp(-A(n+1)^\alpha) + \sum_{m=1}^\infty (2m+1)^\lambda \exp(-A(2m+1)^\alpha(n+1)^\alpha) \right) \\
&\leq 2(n+1)^{\lambda+1} \\
&\times \left( \exp(-A(n+1)^\alpha) + \int_0^\infty (2y+1)^\lambda \exp(-A(2y+1)^\alpha(n+1)^\alpha) dy \right) \\
&\leq 2(1+Cn^{-\alpha})(n+1)^{\lambda+1} \exp(-A(n+1)^\alpha) \\
&\leq C(n+1)^{\lambda+1} \exp(-A(n+1)^\alpha). \tag{2.5}
\end{aligned}$$

### 3. LOWER ESTIMATES

Efficient lower estimates for  $E_n(f)_{C_{2\pi}^*}$ , where  $f$  is an odd or even function with monotone Fourier coefficients, are given in the following lemma.

**Lemma 3.1.** (a) For  $n \geq 1$  and  $f(x) = a_0/2 + \sum_{k=1}^\infty a_k \cos kx \in C_{2\pi}^*$ , where  $\{a_k\}_{k=n+1}^\infty$  is a positive non-increasing sequence, the following inequality holds:

$$E_n(f)_{C_{2\pi}^*} \geq (1/4) \sup_{N \in \mathbb{N}} (N+1)a_{N+n}. \tag{3.1}$$

(b) For  $n \geq 1$  and  $f(x) = \sum_{k=1}^\infty b_k \sin kx \in C_{2\pi}^*$ , where  $\{b_k\}_{k=n+1}^\infty$  is a positive non-increasing sequence, the following inequality holds:

$$E_n(f)_{C_{2\pi}^*} \geq (1/4) \sup_{N \in \mathbb{N}} (N+1)b_{N+n}. \tag{3.2}$$

Inequalities (3.1) and (3.2) follow from more general estimates obtained by Newman and Rivlin [7] and the author [3], respectively. Lower estimates for other classes of functions were developed in [4].

Lower estimates for  $E_n(f)_{C_{2\pi}^*}$ , where a function  $f$  of bounded variation satisfies some additional conditions, are based on the following theorem, which is interesting in itself.

**Theorem 3.1.** Let  $f \in C_{2\pi}^*$  be an absolutely continuous function with  $f' \in L_{2\pi}^*$ . In addition, we assume that if  $Q_n \in \mathcal{T}_n$  is a polynomial of best mean approximation to  $f'$ , then  $f' - Q_n \neq 0$  a. e. on  $\mathbb{R}$  and there exists  $\beta \in [0, 2\pi)$  such that  $f' - Q_n$  has exactly  $2(n+1)$  sign changes on  $[\beta, 2\pi + \beta)$ . Then

$$E_n(f)_{C_{2\pi}^*} \geq [4(n+1)]^{-1} E_n(f')_{L_{2\pi}^*}. \tag{3.3}$$

Equality in (3.3) holds for the function

$$f_n(x) := -\frac{4}{\pi(n+1)} \sum_{m=0}^\infty \frac{\cos[(2m+1)(n+1)x]}{(2m+1)^2}. \tag{3.4}$$

*Proof.* Let  $Q_n \in \mathcal{T}_n$  be a polynomial of best approximation to  $f'$  in  $L_{2\pi}^*$  and let  $x_1, \dots, x_{2(n+1)}$  be the points from  $(\beta, 2\pi + \beta)$ , in which  $f' - Q_n$  changes its sign. Setting  $x_0 := \beta$ ,  $x_{2(n+1)+1} := 2\pi + \beta$  and using the criterion of best approximation

of  $f'$  in  $L_{2\pi}^*$  [8, Section 2.8.1], we have for any polynomial  $T_n \in \mathcal{T}_n$

$$\begin{aligned} 0 &= \int_{\beta}^{2\pi+\beta} T_n(x) \operatorname{sgn}(f'(x) - Q_n(x)) dx = \pm \sum_{i=0}^{2(n+1)} (-1)^i \int_{x_i}^{x_{i+1}} T_n(x) dx \\ &= \pm \left( 2 \sum_{i=0}^{2n+1} (-1)^{i+1} \int_{\beta}^{x_{i+1}} T_n(x) dx - \int_{\beta}^{2\pi+\beta} T_n(x) dx \right). \end{aligned} \quad (3.5)$$

Next, the following trivial relation holds for any constant  $C$ :

$$\sum_{i=0}^{2n+1} (-1)^{i+1} C = 0. \quad (3.6)$$

Further applying (3.5) and (3.6) to  $T_n(x) = \sin kx$  or  $T_n(x) = \cos kx$ ,  $1 \leq k \leq n$ , we obtain that for any  $T_n^*(x) = \sum_{k=1}^n (a_k \cos kx + b_k \sin kx)$ ,

$$\sum_{i=0}^{2n+1} (-1)^{i+1} T_n^*(x_{i+1}) = 0. \quad (3.7)$$

It follows from (3.6) that (3.7) is valid for any  $T_n^* \in \mathcal{T}_n$ . Let us define a measure  $\mu$  with its support on  $\{x_1, \dots, x_{2(n+1)}\}$  by

$$\mu(x_{i+1}) = 2(-1)^{i+1}, \quad 0 \leq i \leq 2n+1. \quad (3.8)$$

Then  $\mu$  is orthogonal to  $\mathcal{T}_n$  since for any  $T_n^* \in \mathcal{T}_n$ , (3.7) is equivalent to the relation

$$\int_{\beta}^{2\pi+\beta} T_n^*(x) d\mu(x) = 0. \quad (3.9)$$

Therefore taking account of the facts that  $f' - Q_n \neq 0$  a. e. on  $\mathbb{R}$  and  $f' - Q_n$  has exactly  $2(n+1)$  sign changes on  $[\beta, 2\pi+\beta]$ , we get from (3.5), (3.6) (3.8), and (3.9) that

$$\begin{aligned} E_n(f')_{L_{2\pi}^*} &= \int_{\beta}^{2\pi+\beta} (f'(x) - Q_n(x)) \operatorname{sgn}(f'(x) - Q_n(x)) dx \\ &= \pm \left( 2 \sum_{i=0}^{2(n+1)} (-1)^{i+1} \int_{\beta}^{x_{i+1}} (f'(x) - Q_n(x)) dx - \int_{\beta}^{2\pi+\beta} (f'(x) - Q_n(x)) dx \right) \\ &= \pm 2 \sum_{i=0}^{2(n+1)} (-1)^{i+1} f(x_{i+1}) = \left| \int_{\beta}^{2\pi+\beta} f(x) d\mu(x) \right| \leq \operatorname{var} \mu E_n(f)_{C_{2\pi}^*}, \end{aligned} \quad (3.10)$$

where

$$\operatorname{var} \mu := \int_{\beta}^{2\pi+\beta} |\mu(x)| = 4(n+1). \quad (3.11)$$

Therefore (3.3) follows from (3.10) and (3.11). To complete the proof, we note that the function  $f_n$  defined by (3.4) is continuous and  $f'_n(x) = \operatorname{sgn} \sin[(n+1)x] \in L_{2\pi}^*$ . Since  $\int_0^{2\pi} f'_n(x) T_n^*(x) dx = 0$  for any  $T_n^* \in \mathcal{T}_n$ , we conclude that  $Q_n(x) = 0$  is a polynomial of best mean approximation to  $f'_n$ . Moreover for  $0 < \beta < \pi/(2n+2)$ , the function  $f'_n - Q_n = f'_n$  has exactly  $2(n+1)$  sign changes on  $(\beta, 2\pi+\beta)$ . Therefore  $f_n$  satisfies all conditions of Theorem 3.1. Finally we have

$$E_n(f'_n)_{L_{2\pi}^*} = \int_0^{2\pi} |f'_n(x)| dx = 2\pi, \quad (3.12)$$

and by Chebyshev's theorem,

$$E_n(f_n)_{C_{2\pi}^*} = \|f_n\|_{C_{2\pi}^*} = \pi/(2n+2), \quad (3.13)$$

since for  $k = 0, 1, \dots, 2n+1$ ,

$$f_n\left(\frac{k\pi}{n+1}\right) = (-1)^{k+1} \frac{4}{\pi(n+1)} \sum_{m=0}^{\infty} \frac{1}{(2m+1)^2} = (-1)^{k+1} \frac{\pi}{2(n+1)}.$$

Thus (3.12) and (3.13) imply

$$E_n(f_n)_{C_{2\pi}^*} = [4(n+1)]^{-1} E_n(f'_n)_{L_{2\pi}^*}.$$

This completes the proof of the theorem.  $\square$

The following corollary shows that estimate (3.3) holds for functions satisfying Nagy's conditions.

**Corollary 3.1.** *If the Fourier coefficients of a function  $g = f' \in L_{2\pi}^*$  satisfy conditions of Lemma 2.1(a) or Lemma 2.1(b), then inequality (3.3) holds.*

*Proof.* Nagy [6] (see also [8, Sections 2.11.5 and 2.13.32]) proved that if conditions of Lemma 1(a) or Lemma 1(b) are satisfied for  $g(x) = f'(x) = \sum_{k=1}^{\infty} B_k \sin kx$  or  $g(x) = f'(x) = A_0/2 + \sum_{k=1}^{\infty} A_k \cos kx$ , then  $\sin(n+1)x(f'(x) - Q_n(x)) \geq 0$  or  $\cos(n+1)x(f'(x) - Q_n(x)) \geq 0$  and  $f' - Q_n \neq 0$  a. e. on  $\mathbb{R}$ . Here  $Q_n$  is a polynomial of best mean approximation to  $f'$ . Therefore the conditions of Theorem 3.1 are satisfied and (3.3) follows.  $\square$

Lower estimates of Lemma 3.1 and Corollary 3.1 are used in the following examples.

**Example 3.1.**  $f(x) = \sum_{k=2}^{\infty} \frac{\sin kx}{k \log^q k}$ ,  $q > 0$ . Then using Lemma 3.1(b), we have

$$E_n(f)_{C_{2\pi}^*} \geq \frac{1}{4} \sup_{N \in \mathbb{N}} \frac{N+1}{(N+n) \log^q(N+n)} \geq \frac{1}{8 \log^q(2n)} \geq \frac{C}{\log^q(n+1)}.$$

**Example 3.2.**  $f(x) = \sum_{k=1}^{\infty} k^\lambda \exp(-A k^\alpha) \cos kx$ ,  $\lambda \in \mathbb{R}$ ,  $A > 0$ ,  $\alpha > 0$ . Then for  $n$  large enough, the Fourier coefficients of  $f'$  satisfy the conditions of Lemma 2.1(b). Therefore by Corollary 3.1,

$$\begin{aligned} E_n(f)_{C_{2\pi}^*} &\geq [4(n+1)]^{-1} E_n(f')_{L_{2\pi}^*} = (1/2)(n+1)^\lambda \\ &\times \left( \exp(-A(n+1)^\alpha) + \sum_{m=0}^{\infty} (2m+1)^\lambda \exp(-A(2m+1)^\alpha(n+1)^\alpha) \right) \\ &\geq C(n+1)^\lambda \exp(-A(n+1)^\alpha). \end{aligned}$$

**Example 3.3.**  $f(x) = \sum_{k=1}^{\infty} k^{-\lambda} \cos kx$ ,  $\lambda > 1$ . Then by Lemma 3.1(a),

$$E_n(f)_{C_{2\pi}^*} \geq \frac{1}{4} \sup_{N \in \mathbb{N}} \frac{N+1}{(N+n)^\lambda} \geq C(n+1)^{-\lambda+1}. \quad (3.14)$$

In the following corollary, we combine upper and lower estimates of Examples 2.1, 2.2, 3.1, and 3.2.



## ON UNIFORM APPROXIMATION OF PERIODIC FUNCTIONS

**Corollary 3.2.** (a) For  $f(x) = \sum_{k=2}^{\infty} \frac{\sin kx}{k \log^q k}$ ,  $q > 0$ ,

$$C_2 \log^{-q}(n+1) \leq E_n(f)_{C_{2\pi}^*} \leq C_1 \log^{-q}(n+1).$$

(b) For  $f(x) = \sum_{k=1}^{\infty} k^\lambda \exp(-A k^\alpha) \cos kx$ ,  $\lambda \in \mathbb{R}$ ,  $A > 0$ ,  $\alpha > 0$ ,

$$\lim_{n \rightarrow \infty} (E_n(f)_{C_{2\pi}^*})^{A^{-1}(n+1)^{-\alpha}} = e^{-1}.$$

## REFERENCES

- [1] N.I. Akhiezer, *Lectures on the Theory of Approximation*, (2nd ed.), Nauka, Moscow, 1965. [Russian]
- [2] V.F. Babenko, S.A. Pichugov, Best linear approximation of some classes of differentiable periodic functions, *Math. Notes*, 27,325-329(1980).
- [3] M.I. Ganzburg, A lower estimate of best approximations of continuous functions, *Ukranian Math. J.*, 41,763-767(1989).
- [4] M.I. Ganzburg, Best approximation of functions like  $|x|^\lambda \exp(-A|x|^{-\alpha})$ , *J. Approx. Theory*, 92,379-410(1998).
- [5] N.P. Korneichuk, *Exact Constants in Approximation Theory*, Cambridge University Press, Cambridge, 1991.
- [6] B. Sz.-Nagy, Über gewisse Extremalfragen bei transformierten trigonometrischen Entwicklungen. I. Periodischer Fall, *Berichte Acad. d. Wiss., Leipzig*, 90,103-134(1938).
- [7] D.J. Newman and T.J. Rivlin, Approximation of monomials by lower degree polynomials, *Aequat. Math.* 14,451-455(1976).
- [8] A.F. Timan, *Theory of Approximation of Functions of a Real Variable*, MacMillan, New York, 1963.
- [9] R.M. Trigub and E.S. Belinsky, *Fourier Analysis and Approximation of Functions*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.
- [10] A. Zygmund, *Trigonometric Series* (2nd ed.), Vol. I, Cambridge University Press, Cambridge, 1959.

DEPARTMENT OF MATHEMATICS, HAMPTON UNIVERSITY, HAMPTON, VIRGINIA 23668  
*E-mail address:* michael.ganzburg@hamptonu.edu

# Some convergence theorems for a class of generalized $\Phi$ -hemicontractive mappings\*

Chang He Xiang, Zhe Chen, Ke Quan Zhao<sup>†</sup>

*College of Mathematics and Computer Science,  
Chongqing Normal University, Chongqing, 400047, P.R. China*

**Abstract.** Suppose that  $E$  is a real normed linear space,  $C$  is a nonempty convex subset of  $E$  and  $T : C \rightarrow C$  is a Lipschitz generalized  $\Phi$ -hemicontractive mapping. Under suitable conditions on the iterative parameters, we show that the Mann iterative sequence with errors converges strongly to the unique fixed point of  $T$ .

**Key Words:**  $\phi$ -hemicontractive mapping; generalized  $\Phi$ -hemicontractive mapping; Mann iterative sequence with errors; fixed point

**2000 Mathematics Subject Classification:** 47H05; 47H10; 54H25

## 1 INTRODUCTION

Let  $E$  be an arbitrary real normed linear space with dual space  $E^*$  and  $C$  be a nonempty subset of  $E$ . We denote by  $J$  the normalized duality mapping from  $E$  to  $2^{E^*}$  defined by

$$J(x) = \{x^* \in E^* : \langle x, x^* \rangle = \|x\|^2 = \|x^*\|^2\}, \forall x \in E,$$

where  $\langle \cdot, \cdot \rangle$  denotes the generalized duality pairing.

A mapping  $T : C \rightarrow E$  is called  $\phi$ -hemicontractive if the fixed point set  $F(T) = \{x \in C : Tx = x\}$  is nonempty and there exists a strictly increasing function  $\phi : [0, \infty) \rightarrow [0, \infty)$  with  $\phi(0) = 0$  such that, for all  $x \in C$  and  $x^* \in F(T)$ , there exists  $j(x - x^*) \in J(x - x^*)$  satisfying

$$\langle Tx - x^*, j(x - x^*) \rangle \leq \|x - x^*\|^2 - \phi(\|x - x^*\|)\|x - x^*\|.$$

$T$  is called *generalized  $\Phi$ -hemicontractive* if the fixed point set  $F(T)$  is nonempty and there exists a strictly increasing function  $\Phi : [0, \infty) \rightarrow [0, \infty)$  with  $\Phi(0) = 0$  such that

$$\langle Tx - x^*, j(x - x^*) \rangle \leq \|x - x^*\|^2 - \Phi(\|x - x^*\|) \quad (1.1)$$

holds for all  $x \in C, x^* \in F(T)$  and for some  $j(x - x^*) \in J(x - x^*)$ .

Generalized  $\Phi$ -hemicontractive mapping is also called uniformly hemicontractive in [1, 2]. It is well known that this kind of mappings play important roles in nonlinear analysis.

---

\*This research is supported by the Education Committee project Research Foundation of Chongqing (Grant No. KJ070806) and Chongqing Key Laboratory of Operations Research and System Engineering.

<sup>†</sup>E-mail address: xch@cqnu.edu.cn (C.H. Xiang), zhe505@yahoo.com.cn (Z. Chen), kequanz@163.com (K.Q. Zhao).

By taking  $\Phi(s) = s\phi(s)$ , where  $\phi : [0, \infty) \rightarrow [0, \infty)$  is a strictly increasing function with  $\phi(0) = 0$ , we know that the class of  $\phi$ -hemicontractive mappings is a subset of the class of generalized  $\Phi$ -hemicontractive mappings. The Example 1.1 below demonstrates that the class of Lipschitz  $\phi$ -hemicontractive mappings is a proper subset of the class of Lipschitz generalized  $\Phi$ -hemicontractive mappings.

**Example 1.1.** Let  $E = \mathbb{R}$  be the reals with the usual norm. Define  $T : E \rightarrow E$  by

$$Tx = x - \frac{x}{1+x^2}, \quad \forall x \in E.$$

Then  $T$  is Lipschitz and  $T$  has the unique fixed point  $x^* = 0 \in E$ . Let  $\Phi : [0, \infty) \rightarrow [0, \infty)$  be defined by  $\Phi(s) = \frac{s^2}{1+s^2}$ . Then  $\Phi(s)$  is a strictly increasing function with  $\Phi(0) = 0$ . For all  $x \in E$ , we have

$$\langle Tx - Tx^*, x - x^* \rangle = \langle Tx, x \rangle = x^2 - \frac{x^2}{1+x^2} = |x|^2 - \Phi(|x|) = |x - x^*|^2 - \Phi(|x - x^*|),$$

so that  $T$  is a Lipschitz generalized  $\Phi$ -hemicontractive mapping. Since

$$\langle Tx - Tx^*, x - x^* \rangle = |x - x^*|^2 - \Phi(|x - x^*|) \leq |x - x^*|^2 - \phi(|x - x^*|)|x - x^*|$$

holds for all  $x \in E$  if and only if  $\phi(s) \leq \frac{s}{1+s^2}$  for all  $s \geq 0$ , we know that  $T$  is not  $\phi$ -hemicontractive since such a function  $\phi$  is not increasing.

Many results have been proved on convergence or stability of Ishikawa iterative sequences (with errors) or Mann iterative sequences (with errors) for Lipschitz  $\phi$ -hemicontractive mappings (see, e.g., [3-8] and the references therein). Recently, there are some authors studied the convergence of the iterative sequences (with errors) involving generalized  $\Phi$ -hemicontractive mappings (see, e.g., [1,2,9-11] and the references therein). In 2005, Chidume and Chidume [11] proved the following results:

**Theorem CC1** [11, Theorem 2.2]. Let  $E$  be a real normed linear space and  $T : E \rightarrow E$  be uniformly continuous. Let  $\{x_n\}$  be a sequence in  $E$  defined iteratively from an arbitrary  $x_0 \in E$  by

$$x_{n+1} = a_n x_n + b_n T x_n + c_n u_n, \quad n \geq 0,$$

where  $\{a_n\}$ ,  $\{b_n\}$ ,  $\{c_n\}$  are sequences in  $[0, 1]$  satisfying the following conditions:

- (i)  $a_n + b_n + c_n = 1, \quad \forall n \geq 0;$
- (ii)  $\sum_{n=0}^{\infty} (b_n + c_n) = \infty;$
- (iii)  $\sum_{n=0}^{\infty} (b_n + c_n)^2 < \infty;$
- (iv)  $\sum_{n=0}^{\infty} c_n < \infty$  and such that

$$\langle Tx_n - x^*, j(x_n - x^*) \rangle \leq \|x_n - x^*\|^2 - \Phi(\|x_n - x^*\|), \quad \forall n \geq 0,$$

where  $\Phi : [0, \infty) \rightarrow [0, \infty)$  is a strictly increasing function with  $\Phi(0) = 0$ , and  $\{u_n\}$  is a bounded sequence in  $E$ . Then  $\{x_n\}$  is bounded.

**Theorem CC2** [11, Theorem 2.3]. Let  $E$  be a real normed linear space,  $K$  be a nonempty subset of  $E$  and  $T : K \rightarrow E$  be a uniformly continuous generalized  $\Phi$ -hemicontractive mapping,

i.e., there exist  $x^* \in F(T)$  and a strictly increasing function  $\Phi : [0, \infty) \rightarrow [0, \infty)$ ,  $\Phi(0) = 0$  such that for all  $x \in K$ , there exists  $j(x - x^*) \in J(x - x^*)$  such that

$$\langle Tx_n - x^*, j(x_n - x^*) \rangle \leq \|x_n - x^*\|^2 - \Phi(\|x_n - x^*\|), \forall n \geq 0.$$

- (a) If  $y^* \in K$  is a fixed point of  $T$ , then  $y^* = x^*$  and so  $T$  has at most one fixed point in  $K$ .  
 (b) Suppose there exists  $x_0 \in K$ , such that the sequence  $\{x_n\}$  defined by

$$x_{n+1} = a_n x_n + b_n T x_n + c_n u_n, \quad n \geq 0,$$

is contained in  $K$ , where  $\{a_n\}$ ,  $\{b_n\}$  and  $\{c_n\}$  are real sequences satisfying the following conditions:

- (i)  $a_n + b_n + c_n = 1$ ;  
 (ii)  $\sum_{n=0}^{\infty} (b_n + c_n) = \infty$ ;  
 (iii)  $\sum_{n=0}^{\infty} (b_n + c_n)^2 < \infty$ ;  
 (iv)  $\sum_{n=0}^{\infty} c_n < \infty$ ; and  $\{u_n\}$  is a bounded sequence in  $E$ .

Then  $\{x_n\}$  converges strongly to  $x^*$ . In particular, if  $y^*$  is a fixed point of  $T$  in  $K$ , then  $\{x_n\}$  converges strongly to  $y^*$ .

The proof of conclusion (b) in Theorem CC2 is based on Theorem CC1. However, there is a gap in the proof of Theorem CC1. In fact, in the proof of Theorem CC1, in order to prove  $\|x_n - x^*\| \leq 2\Phi^{-1}(a_0)$  for all  $n \geq 0$  by induction, where  $a_0$  is a positive number, by assuming that  $\|x_n - x^*\| \leq 2\Phi^{-1}(a_0)$  and  $\|x_{n+1} - x^*\| > 2\Phi^{-1}(a_0)$  hold for some  $n$ , the authors of [11] established following inequality (see [11, page 551])

$$\|x_{n+1} - x^*\|^2 \leq \|x_n - x^*\|^2 - \alpha_n \Phi(2\Phi^{-1}(a_0)) + M_1 \alpha_n^2 + c_n \rho \quad (1.2)$$

for the same  $n$ . Unfortunately, (1.2) does not imply that

$$\Phi(2\Phi^{-1}(a_0)) \sum_{j=0}^n \alpha_j \leq \sum_{n=0}^n \left( \|x_j - x^*\|^2 - \|x_{j+1} - x^*\|^2 \right) + M_1 \sum_{j=0}^{\infty} \alpha_j^2 + \rho \sum_{j=0}^n c_j,$$

since (1.2) holds only for one given natural number  $n$ . Therefore, the result of Theorem CC2 may be not true since its proof is based on Theorem CC1.

On the other hand, it has been proved in [6] that if  $T : K \rightarrow K$  is a Lipschitz  $\phi$ -hemicontractive mapping and  $c_n = 0$  for all  $n \geq 0$ , then the conclusion (b) in Theorem CC2 holds. Since the class of Lipschitz  $\phi$ -hemicontractive mappings is a proper subset of the class of Lipschitz generalized  $\Phi$ -hemicontractive mappings, this leads to the following questions.

**Question 1.** Suppose that  $T : C \rightarrow C$  is a Lipschitz generalized  $\Phi$ -hemicontractive mapping. Does the result in [6] hold?

**Question 2.** Suppose that  $T : C \rightarrow C$  is a Lipschitz generalized  $\Phi$ -hemicontractive mapping. Does the conclusion (b) of Theorem CC2 hold?

The main purpose of this paper is to give affirmative answer to Questions 1 and 2.

## 2 PRELIMINARIES

The following lemmas will be used in the proof of our main results.

**Lemma 2.1** (See, e.g., [11]). Let  $E$  be a real normed linear space. Then for all  $x, y \in E$ , we have

$$\|x + y\|^2 \leq \|x\|^2 + 2\langle y, j(x + y) \rangle, \quad \forall j(x + y) \in J(x + y).$$

**Lemma 2.2** (See, e.g., [12]). Let  $\{a_n\}, \{b_n\}, \{c_n\}$  be three nonnegative sequences satisfying the following condition:

$$a_{n+1} \leq (1 + b_n)a_n + c_n, \quad \forall n \geq n_0,$$

where  $n_0$  is some nonnegative integer,  $\sum_{n=n_0}^{\infty} b_n < \infty$  and  $\sum_{n=n_0}^{\infty} c_n < \infty$ . Then the limit  $\lim_{n \rightarrow \infty} a_n$  exists.

**Lemma 2.3.** Suppose that there exists a natural number  $n_0$  such that  $a_n, b_n, c_n$  and  $\beta_n$  are nonnegative real numbers for all  $n \geq n_0$  satisfying the following conditions:

- (i)  $a_{n+1} \leq (1 + b_n)a_n - \beta_n \varphi(a_{n+1}) + c_n, \quad \forall n \geq n_0,$
- (ii)  $\sum_{n=n_0}^{\infty} b_n < \infty, \quad \sum_{n=n_0}^{\infty} c_n < \infty,$
- (iii)  $\sum_{n=n_0}^{\infty} \beta_n = \infty,$

where  $\varphi : [0, \infty) \rightarrow [0, \infty)$  is a strictly increasing function with  $\varphi(0) = 0$ . Then  $\lim_{n \rightarrow \infty} a_n = 0$ .

*Proof.* By condition (i), we have

$$a_{n+1} \leq (1 + b_n)a_n + c_n, \quad \forall n \geq n_0.$$

Using condition (ii) and Lemma 2.2, we obtain that  $\lim_{n \rightarrow \infty} a_n$  exists and so  $\{a_n\}$  is bounded. Suppose that  $\lim_{n \rightarrow \infty} a_n = a$  and  $a_n \leq M$  ( $\forall n \geq n_0$ ), where  $a, M$  are nonnegative constants.

Let  $d_n = Mb_n + c_n$ . Then  $\sum_{n=n_0}^{\infty} d_n < \infty$  by condition (ii). It follows from condition (i) that

$$a_{n+1} \leq a_n - \beta_n \varphi(a_{n+1}) + d_n \quad (\forall n \geq n_0). \quad (2.1)$$

Now, we prove that  $a = 0$ . If  $a > 0$ , then there exists a nonnegative integer  $N \geq n_0$  such that  $a_{n+1} \geq \frac{a}{2}$  for all  $n \geq N$ . Since  $\varphi$  is strictly increasing, we have  $\varphi(a_{n+1}) \geq \varphi(\frac{a}{2}) > 0$  for all  $n \geq N$ . It follows from (2.1) that  $a_{n+1} \leq a_n - \beta_n \varphi(\frac{a}{2}) + d_n$  ( $\forall n \geq N$ ) and so

$$\infty = \varphi(\frac{a}{2}) \sum_{n=N}^{\infty} \beta_n \leq a_N + \sum_{n=N}^{\infty} d_n < \infty,$$

which is a contradiction. Therefore,  $\lim_{n \rightarrow \infty} a_n = a = 0$ . This completes the proof.  $\square$

**Remark 2.1.** Lemma 2.3 is different from Lemma 3 in [13], which requires that  $b_n = 0$  for all  $n \geq 0$  and  $c_n = o(\beta_n)$ .

### 3 MAIN RESULTS

**Theorem 3.1.** Let  $E$  be a real normed linear space,  $C$  be a nonempty convex subset of  $E$  and  $T : C \rightarrow C$  be a Lipschitz generalized  $\Phi$ -hemicontractive mapping. For given  $x_0 \in C$ , suppose that the sequence  $\{x_n\} \subset C$  is the Mann iterative sequence with errors defined by

$$x_{n+1} = \alpha_n x_n + \beta_n T x_n + \gamma_n u_n, \quad n \geq 0, \quad (3.1)$$

where  $\{u_n\}$  is a bounded sequence in  $C$  and  $\{\alpha_n\}, \{\beta_n\}, \{\gamma_n\}$  are sequences in  $[0, 1]$  satisfying the following conditions:

- (1)  $\alpha_n + \beta_n + \gamma_n = 1 (\forall n \geq 0)$ ;
- (2)  $\sum_{n=0}^{\infty} \beta_n = \infty$ ;
- (3)  $\sum_{n=0}^{\infty} \beta_n^2 < \infty, \sum_{n=0}^{\infty} \gamma_n < \infty$ .

Then  $\{x_n\}$  converges strongly to the unique fixed point of  $T$  in  $C$ .

*Proof.* It follows from (1.1) that  $F(T) = \{x \in C : Tx = x\}$  is singleton. Let  $F(T) = \{p\}$  and  $M = \sup\{\|u_n - p\| : n \geq 0\}$ . Then  $M < \infty$  and

$$\|u_n - x_n\| \leq \|u_n - p\| + \|p - x_n\| \leq M + \|x_n - p\|. \quad (3.2)$$

By (1.1), there exists  $j(x_n - p) \in J(x_n - p)$  such that

$$\langle Tx_n - p, j(x_n - p) \rangle \leq \|x_n - p\|^2 - \Phi(\|x_n - p\|), \quad \forall n \geq 0, \quad (3.3)$$

where  $\Phi : [0, \infty) \rightarrow [0, \infty)$  is a strictly increasing function with  $\Phi(0) = 0$ . Let  $L$  be the Lipschitz constant of  $T$ . By Lemma 1.1, (3.1) and (3.3), we obtain

$$\begin{aligned} \|x_{n+1} - p\|^2 &= \|\alpha_n(x_n - p) + \beta_n(Tx_n - p) + \gamma_n(u_n - p)\|^2 \\ &\leq \alpha_n^2 \|x_n - p\|^2 + 2\beta_n \langle Tx_n - p, j(x_{n+1} - p) \rangle \\ &\quad + 2\gamma_n \langle u_n - p, j(x_{n+1} - p) \rangle \\ &\leq \alpha_n^2 \|x_n - p\|^2 + 2\beta_n \langle Tx_{n+1} - p, j(x_{n+1} - p) \rangle \\ &\quad + 2\beta_n \langle Tx_n - Tx_{n+1}, j(x_{n+1} - p) \rangle + 2\gamma_n M \|x_{n+1} - p\| \\ &\leq \alpha_n^2 \|x_n - p\|^2 + 2\beta_n [\|x_{n+1} - p\|^2 - \Phi(\|x_{n+1} - p\|)] \\ &\quad + 2\beta_n L \|x_n - x_{n+1}\| \cdot \|x_{n+1} - p\| + 2\gamma_n M \|x_{n+1} - p\|, \quad \forall n \geq 0. \end{aligned} \quad (3.4)$$

From (3.1) and condition (1), we have

$$x_n - x_{n+1} = \beta_n(x_n - Tx_n) - \gamma_n(u_n - x_n).$$

It follows from (3.2) that

$$\|x_n - x_{n+1}\| \leq \beta_n \|x_n - Tx_n\| + \gamma_n (M + \|x_n - p\|). \quad (3.5)$$

Observe that

$$\|x_n - Tx_n\| \leq \|x_n - p\| + \|p - Tx_n\| \leq (1 + L)\|x_n - p\|. \quad (3.6)$$

Taking (3.6) into (3.5), we obtain

$$\|x_n - x_{n+1}\| \leq [(1 + L)\beta_n + \gamma_n] \|x_n - p\| + M\gamma_n. \quad (3.7)$$

Taking (3.7) into (3.4), we have

$$\begin{aligned} \|x_{n+1} - p\|^2 &\leq \alpha_n^2 \|x_n - p\|^2 + 2\beta_n [\|x_{n+1} - p\|^2 - \Phi(\|x_{n+1} - p\|)] \\ &\quad + 2(\delta_n \|x_n - p\| + \sigma_n) \|x_{n+1} - p\|, \end{aligned} \quad (3.8)$$

where  $\delta_n = L\beta_n [(1 + L)\beta_n + \gamma_n]$ ,  $\sigma_n = (L\beta_n + 1)M\gamma_n$ ,  $\forall n \geq 0$ . By condition (3), we have

$$\sum_{n=0}^{\infty} \delta_n < \infty, \quad \sum_{n=0}^{\infty} \sigma_n < \infty. \quad (3.9)$$

Denote  $a_n = \|x_n - p\|^2$  ( $\forall n \geq 0$ ) and  $\varphi(s) = 2\Phi(\sqrt{s})$ . It follows from (3.8) that

$$a_{n+1} \leq \alpha_n^2 a_n + 2\beta_n a_{n+1} - \beta_n \varphi(a_{n+1}) + 2(\delta_n \sqrt{a_n} + \sigma_n) \sqrt{a_{n+1}}, \quad \forall n \geq 0.$$

Noting that  $0 \leq \alpha_n \leq 1 - \beta_n$ , we obtain

$$\begin{aligned} a_{n+1} &\leq (1 - \beta_n)^2 a_n + 2\beta_n a_{n+1} - \beta_n \varphi(a_{n+1}) + \delta_n (a_n + a_{n+1}) + \sigma_n (1 + a_{n+1}) \\ &= (1 - 2\beta_n + \beta_n^2 + \delta_n) a_n + (2\beta_n + \delta_n + \sigma_n) a_{n+1} - \beta_n \varphi(a_{n+1}) + \sigma_n. \end{aligned} \quad (3.10)$$

It follows from (3.9) and condition (3) that  $\lim_{n \rightarrow \infty} (2\beta_n + \delta_n + \sigma_n) = 0$ . Thus, there exists a natural number  $n_0$  such that  $2\beta_n + \delta_n + \sigma_n \leq \frac{1}{2}$  for all  $n \geq n_0$ . Let

$$\begin{aligned} b_n &= \frac{1 - 2\beta_n + \beta_n^2 + \delta_n}{1 - 2\beta_n - \delta_n - \sigma_n} - 1 = \frac{\beta_n^2 + 2\delta_n + \sigma_n}{1 - 2\beta_n - \delta_n - \sigma_n}, \\ c_n &= \frac{\sigma_n}{1 - 2\beta_n - \delta_n - \sigma_n}. \end{aligned}$$

By (3.10),

$$a_{n+1} \leq (1 + b_n) a_n - \beta_n \varphi(a_{n+1}) + c_n, \quad \forall n \geq n_0.$$

Since  $\frac{1}{2} \leq 1 - 2\beta_n - \delta_n - \sigma_n \leq 1$  for all  $n \geq n_0$ ,

$$0 \leq b_n \leq 2(\beta_n^2 + 2\delta_n + \sigma_n), \quad 0 \leq c_n \leq 2\sigma_n, \quad \forall n \geq n_0.$$

It follows from (3.9) and condition (3) that  $\sum_{n=n_0}^{\infty} b_n < \infty$  and  $\sum_{n=n_0}^{\infty} c_n < \infty$ . Therefore, by using Lemma 2.3 and condition (2), we obtain that  $\lim_{n \rightarrow \infty} \|x_n - p\|^2 = \lim_{n \rightarrow \infty} a_n = 0$ . That is,  $\lim_{n \rightarrow \infty} \|x_n - p\| = 0$ . This completes the proof of Theorem 3.1.  $\square$

**Remark 3.1.** Theorem 3.1 gives an affirmative answer to Question 2.

Taking  $\gamma_n = 0$  for all  $n \geq 0$  in Theorem 3.1, we have the following result.

**Theorem 3.2.** Let  $E$  be a real normed linear space,  $C$  be a nonempty convex subset of  $E$  and  $T : C \rightarrow C$  be a Lipschitz generalized  $\Phi$ -hemicontractive mapping. For given  $x_0 \in C$ , suppose that the sequence  $\{x_n\} \subset C$  is the Mann iterative sequence defined by

$$x_{n+1} = (1 - \beta_n)x_n + \beta_n T x_n, \quad n \geq 0,$$

where  $\{\beta_n\}$  is a sequence in  $[0, 1]$  satisfying the following conditions:

- (1)  $\sum_{n=0}^{\infty} \beta_n = \infty$ ;
- (2)  $\sum_{n=0}^{\infty} \beta_n^2 < \infty$ .

Then  $\{x_n\}$  converges strongly to the unique fixed point of  $T$  in  $C$ .

**Remark 3.2.** Theorem 3.2 gives an affirmative answer to Question 1.

## References

- [1] C. Moore and B.V.C. Nnoli, Iterative solution of nonlinear equations involving set-valued uniformly accretive operators, *Comput. Math. Appl.* **42**(2001), 131-140.
- [2] C.E. Chidume and H. Zegeye, Approximation methods for nonlinear operator equations, *Proc. Amer. Math. Soc.* **131**(2003), 2467-2478.
- [3] M.O. Osilike, Iterative solution of nonlinear equations of the  $\phi$ -strongly accretive type, *J. Math. Anal. Appl.* **200**(1996), 259-271.
- [4] Y.G. Xu, Ishikawa and Mann iterative processes with errors for nonlinear strongly accretive operator equations, *J. Math. Anal. Appl.* **224**(1998), 91-101.
- [5] X.P. Ding, Iterative process with errors to nonlinear  $\phi$ -strongly accretive operator equations in arbitrary Banach spaces, *Comput. Math. Appl.* **33**(1997), 75-82.
- [6] M.O. Osilike, Iterative solution of nonlinear  $\phi$ -strongly accretive operator equations in arbitrary Banach spaces, *Nonlinear Anal.* **36**(1999), 1-9.
- [7] R.P. Agarwal, N.J. Huang and Y.J. Cho, Stability of iterative processes with errors for nonlinear equations of  $\phi$ -strongly accretive type operators, *Numer. Funct. Anal. Optimiz.* **22**(2001), 471-485.
- [8] Z.Y. Huang, Weak stability of Mann and Ishikawa iterations with errors for  $\phi$ -hemiccontractive operators, *Appl. Math. Lett.* **20**(2007), 470-475.
- [9] S.S. Zhang, On the convergence problems of Ishikawa and Mann iteration process with errors for  $\Phi$ -pseudo contractive type mappings, *Appl. Math. Mech.* **21**(2000), 1-12.
- [10] F. Gu, Convergence theorems for  $\Phi$ -pseudocontractive type mappings in normed linear spaces, *Northeast Math. J.* **17**(2001), 340-346.
- [11] C.E. Chidume and C.O. Chidume, Convergence theorems for fixed points of uniformly continuous generalized  $\Phi$ -hemi-contractive mappings, *J. Math. Anal. Appl.* **303** (2005), 545-554.
- [12] S.S. Chang, K.K. Tan, H.W.J. Lee and C.K. Chan, On the convergence of implicit iteration process with error for a finite family of asymptotically nonexpansive mappings, *J. Math. Anal. Appl.* **313** (2006), 273-283.
- [13] C.E. Chidume and E.U. Ofoedu, A new iteration process for generalized Lipschitz pseudo-contractive and generalized Lipschitz accretive mappings, *Nonlinear Anal.* **67** (2007), 307-315.



# ITERATIVE ALGORITHMS FOR A COUNTABLE FAMILY OF NONEXPANSIVE MAPPINGS

YISHENG SONG AND XIAO LIU  
COLLEGE OF MATHEMATICS AND INFORMATION SCIENCE,  
HENAN NORMAL UNIVERSITY, P.R. CHINA, 453007.

**ABSTRACT.** In this paper, strong convergence of Modified Mann type iteration is used to find some common fixed point of a countable family  $\{T_n\}_{n=1}^{+\infty}$  of nonexpansive mappings in Banach space, and their proof is different from ones of Aoyama et al. [Approximation of common fixed points of a countable family of nonexpansive mappings in a Banach space, *Nonlinear Analysis*, 67(2007) 2350–2360] and other existing results which is independent of the convergence of the implicit anchor-like continuous path  $z_t$ , defined by  $z_t = tu + (1-t)Tz_t$ .

**Key Words and Phrases:** Countable family of nonexpansive mappings; Modified Mann type iteration; Mann type iteration; uniformly Gâteaux differentiable.

## 1. Introduction

Let  $E$  be a Banach space and  $E^*$  be its dual space. Let  $K$  be a nonempty closed convex subset of  $E$  and  $T : K \rightarrow K$  be a mapping.  $T$  is said to be non-expansive if  $\|Tx - Ty\| \leq \|x - y\|$  for all  $x, y \in K$ .

In order to find a fixed point of nonexpansive mapping  $T$ , Mann [14] and Halpern [7] respectively introduced the iteration procedure in a Hilbert space as follows (We refer them to as Mann iteration and Halpern iteration):

$$x_{n+1} = (1 - \alpha_n)Tx_n + \alpha_n x_n \quad (1.1)$$

and

$$x_{n+1} = (1 - \alpha_n)Tx_n + \alpha_n u, \quad (1.2)$$

where  $\{\alpha_n\}$  is a sequences in  $[0, 1]$ . Subsequently, Mann iteration and Halpern iteration were studied extensively over the last twenty years for constructions of fixed points of nonlinear mappings and of solutions of nonlinear operator equations involving monotone, accretive and pseudocontractive operators. For example, [3, 4, 11, 12, 13, 16, 19, 20, 21, 26, 27, 28, 31, 32, 33, 35, 36] and many other results which isn't mentioned here.

The modified version of Mann iteration and Halpern iteration were investigated widely by many mathematic workers. For example, Kim-Xu [10] and Chidume-Chidume [5] dealt with the strong convergence of the following iterative scheme (so-called Modified Mann iteration) for a non-expansive mapping  $T$ : for  $x_0, u \in K$ ,

$$x_{n+1} = \alpha_n u + (1 - \alpha_n)(\beta_n x_n + (1 - \beta_n)Tx_n). \quad (1.3)$$

---

1991 *Mathematics Subject Classification.* 47H06, 47J05, 47J25, 47H10, 47H17.  
Email: songyisheng123@yahoo.com.cn, xliu@spaceweather.ac.cn.

where  $\alpha_n, \beta_n \in [0, 1]$ . Song-Chen [25] researched strong convergence of Modified Mann iteration (1.3) in the frame of reflexive Banach space which is complementary and development of the above results.

Recently, for a nonexpansive mappings sequence  $\{T_n\}_{n=1}^{+\infty}$  with some special condition, Jung [9] and O'Hara et al. [17, 18] respectively studied strong convergence of the following iteration: for  $x_0, u \in K$ ,

$$x_{n+1} = \alpha_n u + (1 - \alpha_n)T_n x_n, \quad (1.4)$$

where  $\alpha_n \in [0, 1]$  such that (C1)  $\lim_{n \rightarrow \infty} \alpha_n = 0$  and (C2)  $\sum_{n=1}^{\infty} \alpha_n = \infty$ . Unfortunately, there was a gap in the proof lines of their main results. With the purpose of overcoming the gap, Song-Chen [29, 30] introduced the conception of a uniformly asymptotically regular for  $\{T_n\}_{n=1}^{+\infty}$  and proved several strong convergence results by using the conception. Other investigation of approximating common fixed point for countable family of nonexpansive mappings by means of the iteration (1.4) can be found in Refs [1, 22, 24] and many results which isn't cited here.

Very recently, still for a nonexpansive mappings sequence  $\{T_n\}_{n=1}^{+\infty}$  with some specific condition, Aoyama et al.[2] introduced Mann type iteration procedure: let  $x_1 \in K$  and

$$x_{n+1} = \alpha_n x_n + (1 - \alpha_n)T_n x_n, \quad (1.5)$$

where  $\alpha_n \in [a, b] \subset [0, 1]$ , and showed its strong and weak convergence in uniformly convex Banach space. At the same time, Song [23] also carefully researched the convergence of the iteration (1.5) by the aid of the uniformly asymptotically regular of  $\{T_n\}_{n=1}^{+\infty}$  in a reflexive Banach space.

In this paper, for a countable family  $\{T_n\}_{n=1}^{+\infty}$  of nonexpansive mappings with some appropriate condition (see section 3), we will introduce the following iteration procedure: let  $x_1, u \in K$  and

$$x_{n+1} = \alpha_n u + \beta_n x_n + (1 - \alpha_n - \beta_n)T_n x_n, \quad (1.6)$$

and show its strong convergence in Banach space when  $\alpha_n, \beta_n \in [0, 1]$  satisfy the conditions  $\lim_{n \rightarrow \infty} \alpha_n = 0$ ,  $\sum_{n=0}^{\infty} \alpha_n = \infty$  and  $0 < \liminf_{n \rightarrow \infty} \beta_n \leq \limsup_{n \rightarrow \infty} \beta_n < 1$ . We also go on exploring the weak convergence of the Mann type iteration (1.5) under the condition  $0 < \liminf_{n \rightarrow \infty} \alpha_n \leq \limsup_{n \rightarrow \infty} \alpha_n < 1$ .

## 2. Preliminaries and basic results

Throughout this paper, the fixed point set of  $T$  is denoted by  $F(T) := \{x \in K; Tx = x\}$ . Let  $E$  be a real Banach space and let  $J$  denote the normalized duality mapping from  $E$  into  $2^{E^*}$  given by

$$J(x) = \{f \in E^*, \langle x, f \rangle = \|x\| \|f\|, \|x\| = \|f\|\}, \forall x \in E,$$

where  $E^*$  is the dual space of  $E$  and  $\langle \cdot, \cdot \rangle$  denotes the generalized duality pairing. We write  $x_n \rightharpoonup x$  (respectively  $x_n \xrightarrow{*} x$ ) to indicate that the sequence  $x_n$  weakly (respectively weak\*) converges to  $x$ ; as usual  $x_n \rightarrow x$  will symbolize strong convergence. In order to show our main results, the following conceptions and lemmas are needed.

Let  $S(E) := \{x \in E; \|x\| = 1\}$  denote the unit sphere of a Banach space  $E$ . A Banach space  $E$  is said to have

(i) a *Gâteaux differentiable norm* (we also say that  $E$  is *smooth*), if the limit

$$\lim_{t \rightarrow 0} \frac{\|x + ty\| - \|x\|}{t} \quad (*)$$

exists for each  $x, y \in S(E)$ ;

(ii) a *uniformly Gâteaux differentiable norm*, if for each  $y$  in  $S(E)$ , the limit  $(*)$  is uniformly attained for  $x \in S(E)$ ;

(iii) a *Fréchet differentiable norm*, if for each  $x \in S(E)$ , the limit  $(*)$  is attained uniformly for  $y \in S(E)$ ;

(iv) a *uniformly Fréchet differentiable norm* (we also say that  $E$  is *uniformly smooth*), if the limit  $(*)$  is attained uniformly for  $(x, y) \in S(E) \times S(E)$ .

(v) *fixed point property* for nonexpansive self-mappings, if each nonexpansive self-mapping defined on any bounded closed convex subset  $K$  of  $E$  has at least a fixed point.

A Banach space  $E$  is said to be (vi) *strictly convex* if

$$\|x\| = \|y\| = 1, x \neq y \text{ implies } \frac{\|x + y\|}{2} < 1;$$

(vii) *uniformly convex* if for all  $\varepsilon \in [0, 2]$ ,  $\exists \delta_\varepsilon > 0$  such that

$$\|x\| = \|y\| = 1 \text{ implies } \frac{\|x + y\|}{2} < 1 - \delta_\varepsilon \text{ whenever } \|x - y\| \geq \varepsilon.$$

(viii) The subset  $K$  of  $E$  is a *Chebyshev set*, if  $\forall x \in E$ , there exactly exists unique element  $y \in K$  such that  $\|x - y\| = d(x, K) = \inf\{\|x - z\|; z \in K\}$ .

The following results is well known which are found in reference[8, 34, 15]: the normalized duality mapping  $J$  in a Banach space  $E$  with a uniformly Gâteaux differentiable norm is single-valued and strong-weak\* uniformly continuous on any bounded subset of  $E$ ; each uniformly convex Banach space  $E$  is reflexive and strictly convex and has fixed point property for nonexpansive self-mappings; every uniformly smooth Banach space  $E$  is a reflexive Banach space with a uniformly Gâteaux differentiable norm and has fixed point property for nonexpansive self-mappings; every nonempty closed convex subset is a Chebyshev set in strictly convex and reflexive Banach space (see [15, Corollary 5.1.19]); Each weakly compact convex subset is a Chebyshev set in strictly convex Banach space  $E$  (see [8, Lemma 9.3.7]).

**Lemma 2.1** ([33, Lemma 2.2]) *Let  $\{x_n\}$  and  $\{y_n\}$  be two bounded sequences in a Banach space  $E$  and  $\beta_n \in [0, 1]$  with  $0 < \liminf_{n \rightarrow \infty} \beta_n \leq \limsup_{n \rightarrow \infty} \beta_n < 1$ . Suppose  $x_{n+1} = \beta_n x_n + (1 - \beta_n)y_n$  for all integers  $n \geq 1$  and*

$$\limsup_{n \rightarrow \infty} (\|y_{n+1} - y_n\| - \|x_{n+1} - x_n\|) \leq 0.$$

*Then  $\lim_{n \rightarrow \infty} \|x_n - y_n\| = 0$ .*

**Lemma 2.2** (see [8, Lemma 9.3.6]) *Let  $C$  be a weakly compact subset in Banach space  $E$  and let  $f : E \rightarrow \mathbb{R}$  be a weakly lower semi-continuous function. Then the infimum of  $f$  is achieved in  $C$ .*

In the proof of our main theorems, we also need the following definitions and results. Let  $\mu$  be a continuous linear functional on  $l^\infty$  satisfying  $\|\mu\| = 1 = \mu(1)$ . Then we know that  $\mu$  is a mean on  $N$  if and only if

$$\inf\{a_n; n \in N\} \leq \mu(a) \leq \sup\{a_n; n \in N\}$$

for every  $a = (a_1, a_2, \dots) \in l^\infty$ . According to time and circumstances, we  $\mu_n(a_n)$  instead of  $\mu(a)$ . A mean  $\mu$  on  $N$  is called a *Banach limit* if

$$\mu_n(a_n) = \mu_n(a_{n+1})$$

for every  $a = (a_1, a_2, \dots) \in l^\infty$ . Furthermore, we know the following results [35, 34].

**Lemma 2.3** ([35, Lemma 1]) *Let  $C$  be a nonempty closed convex subset of Banach space  $E$  with uniformly Gâteaux differentiable norm. Let  $\{x_n\}$  be a bounded sequence of  $E$  and let  $\mu_n$  be a mean  $\mu$  on  $\mathbb{N}$  and  $z \in C$ . Then*

$$\mu_n \|x_n - z\|^2 = \min_{y \in C} \mu_n \|x_n - y\|^2$$

*if and only if*

$$\mu_n \langle y - z, J(x_n - z) \rangle \leq 0, \quad \forall y \in C.$$

**Lemma 2.4** ([32, Proposition 2]) *Let  $\alpha$  is a real number and  $(x_0, x_1, \dots) \in l^\infty$  such that  $\mu_n x_n \leq \alpha$  for all Banach Limits. If  $\limsup_{n \rightarrow \infty} (x_{n+1} - x_n) \leq 0$ , then  $\limsup_{n \rightarrow \infty} x_n \leq \alpha$ .*

**Lemma 2.5** ([36]) *Let  $\{a_n\}$  be a sequence of nonnegative real numbers satisfying the property*

$$a_{n+1} \leq (1 - \gamma_n)a_n + \gamma_n \beta_n, \quad n \geq 0,$$

*where  $\{\gamma_n\} \subset (0, 1)$  and  $\{\beta_n\} \subset \mathbb{R}$  such that*

$$(i) \sum_{n=0}^{\infty} \gamma_n = \infty; \quad (ii) \limsup_{n \rightarrow \infty} \beta_n \leq 0.$$

*Then  $\{a_n\}$  converges to zero, as  $n \rightarrow \infty$ .*

### 3. Strong convergence of the modified Mann type iteration

Let  $K$  be a nonempty closed convex subset of Banach space  $E$ . Suppose  $\{T_n\}$  ( $n = 1, 2, \dots$ ) is a countable family of nonexpansive mappings from  $K$  into itself such that  $F := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$ . Recently, in uniformly convex Banach space, Aoyama et al.[1] obtained the strong convergence of modified Halpern type iteration (1.4) if  $\{T_n\}_{n=1}^{+\infty}$  and  $\{\alpha_n\}_{n=1}^{+\infty} \subset (0, 1]$  satisfy the following conditions:

$$(B1) \sum_{n=0}^{\infty} \sup_{x \in C} \|T_{n+1}x - T_nx\| < +\infty \text{ for any bounded subset } C \text{ of } K;$$

$$(C1) \lim_{n \rightarrow \infty} \alpha_n = 0 \text{ and } \sum_{n=0}^{\infty} \alpha_n = \infty;$$

$$(C2) \text{ either } \lim_{n \rightarrow \infty} \frac{\alpha_{n+1}}{\alpha_n} = 1 \text{ or } \sum_{n=0}^{\infty} |\alpha_{n+1} - \alpha_n| < +\infty.$$

Their proof also depend on the following important fact((B1) implies (B2), see [1, Lemma 3.2]):

(B2) for any bounded subset  $C$  of  $K$ , there exists a nonexpansive mapping  $T$  of  $K$  into itself such that

$$\lim_{n \rightarrow \infty} \sup_{x \in C} \|Tx - T_nx\| = 0 \text{ and } F(T) = F.$$

## COUNTABLE FAMILY OF NONEXPANSIVE MAPPINGS

In this section, we introduce the following modified Mann type iteration: for  $x_1, u \in K$ ,

$$x_{n+1} = \alpha_n u + \beta_n x_n + (1 - \alpha_n - \beta_n) T_n x_n, \quad (3.1)$$

and show its strong convergence under the conditions (C1) and (B2) along with

$$(C3) \quad 0 < \liminf_{n \rightarrow \infty} \beta_n \leq \limsup_{n \rightarrow \infty} \beta_n < 1.$$

With the purpose of proving main results, we first show the following lemma which doesn't depend on the convergence of the implicit anchor-like continuous path  $z_t = tu + (1-t)Tz_t$  as many existent results (see [5, 7, 10, 32, 35, 36] and so on).

**Lemma 3.1** *Let  $K$  be either a nonempty weakly compact convex subset of a strictly convex Banach space  $E$  or a nonempty closed convex subset of a reflexive Banach space  $E$  with fixed point property for nonexpansive self-mappings. Assume that  $T : K \rightarrow K$  is a nonexpansive mapping with  $F(T) \neq \emptyset$  and a bounded sequence  $\{x_n\}$  of  $K$  satisfies*

$$\lim_{n \rightarrow \infty} \|x_{n+1} - Tx_n\| = 0 \text{ and } \lim_{n \rightarrow \infty} \|x_n - x_{n+1}\| = 0.$$

*Suppose that  $E$  has a uniformly Gâteaux differentiable norm. Then there exists  $x^* \in F(T)$  such that*

$$\limsup_{n \rightarrow \infty} \langle u - x^*, J(x_{n+1} - x^*) \rangle \leq 0 \text{ for each } u \in K.$$

**Proof.** Let

$$g(x) = \mu_n \|x_n - x\|^2, \forall x \in K.$$

Then  $g(x)$  is continuous and convex on  $K$ . Define a set

$$K_1 = \{x \in K; g(x) = \inf_{y \in K} g(y)\}.$$

From Lemma 2.2 or the reflexivity of  $E$  and the property of  $g(x)$  together with the boundedness of  $\{x_n\}$ , we obtain  $K_1$  is a nonempty bounded closed convex subset of  $K$  and hence weak compact.

For  $\forall x \in K_1$ , then

$$\begin{aligned} g(Tx) &= \mu_n \|x_n - Tx\|^2 = \mu_n \|x_{n+1} - Tx\|^2 \\ &\leq \mu_n (\|x_{n+1} - Tx_n\| + \|Tx_n - Tx\|)^2 \\ &\leq \mu_n \|x_n - x\|^2 = g(x). \end{aligned}$$

Hence,  $Tx \in K_1$ . Namely,  $T(K_1) \subset K_1$ .

**Case 1.** Assumed that  $E$  is strictly convex. Taking  $y \in F(T)$ , then there exists unique  $x^* \in K_1$  such that

$$\|y - x^*\| = d(y, K_1) = \inf_{x \in K_1} \|y - x\|.$$

By  $Tx^* \in K_1$ , we have

$$\|y - Tx^*\| = \|Ty - Tx^*\| \leq \|y - x^*\|.$$

Hence  $x^* = Tx^*$  by the uniqueness of  $x^*$  in  $K_1$ .

**Case 2.** Assumed that  $E$  has fixed point property for nonexpansive self-mappings. By  $T(K_1) \subset K_1$ , there exists  $x^* \in K_1$  such that  $x^* = Tx^*$ .

Using Lemma 2.3 and the definition of  $K_1$ , we get that for  $u \in K$ ,

$$\mu_n \langle u - x^*, J(x_n - x^*) \rangle \leq 0.$$

On the other hand, as  $\lim_{n \rightarrow \infty} \|x_{n+1} - x_n\| = 0$  together with the norm-weak\* uniformly continuity of the duality mapping  $J$  in Banach space with a uniformly Gâteaux differentiable norm, we have

$$\lim_{n \rightarrow \infty} (\langle u - x^*, J(x_{n+1} - x^*) \rangle - \langle u - x^*, J(x_n - x^*) \rangle) = 0.$$

Hence, the sequence  $\{\langle u - x^*, J(x_n - x^*) \rangle\}$  satisfies the conditions of Lemma 3.4. As a result, we must have

$$\limsup_{n \rightarrow \infty} \langle u - x^*, J(x_{n+1} - x^*) \rangle \leq 0.$$

**Theorem 3.2** *Let  $E$  be a reflexive Banach space with a uniformly Gâteaux differentiable norm and having fixed point property for nonexpansive self-mappings. Assume that  $K$  is a nonempty closed convex subset of  $E$  and  $\{T_n\}_{n=1}^{+\infty}$  is a countable family of nonexpansive mappings from  $K$  into itself such that  $F := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$  and the condition (B2). Let  $\{\alpha_n\}$  and  $\{\beta_n\}$  be two real number sequences in  $[0, 1]$  satisfying (C1) and (C3), respectively. Then the modified Mann type iteration sequence  $\{x_n\}$ , defined by (3.1) strongly converges to some point of  $F$ .*

**Proof.** At first, we show that  $\{x_n\}$  is bounded. Taking  $p \in F(T)$ , we have

$$\begin{aligned} \|x_{n+1} - p\| &\leq (1 - \alpha_n - \beta_n)\|T_n x_n - p\| + \beta_n\|x_n - p\| + \alpha_n\|u - p\| \\ &\leq (1 - \alpha_n - \beta_n)\|x_n - p\| + \beta_n\|x_n - p\| + \alpha_n\|u - p\| \\ &\leq \max\{\|x_n - p\|, \|u - p\|\} \\ &\vdots \\ &\leq \max\{\|x_0 - p\|, \|u - p\|\}. \end{aligned}$$

Thus,  $\{x_n\}$  is bounded, and hence so is  $\{T_n x_n\}$  by  $\|T_n x_n - p\| \leq \|x_n - p\|$ .

Next we prove that

$$\lim_{n \rightarrow \infty} \|x_{n+1} - T x_n\| = 0 \text{ and } \lim_{n \rightarrow \infty} \|x_n - x_{n+1}\| = 0. \quad (3.2)$$

Indeed, let  $\lambda_n = \frac{\alpha_n}{1 - \beta_n}$  and  $z_n = \lambda_n u + (1 - \lambda_n)T_n x_n$ . Then

$$\lim_{n \rightarrow \infty} \lambda_n = 0 \text{ and } x_{n+1} = \beta_n x_n + (1 - \beta_n)z_n. \quad (3.3)$$

Therefore, for some constant  $M$  such that  $M > \max\{\|u\|, \sup_{n \in \mathbb{N}} \|T_n x_n\|\}$  and any bounded subset  $C$  of  $K$  containing  $\{x_n\}$ , we have

$$\begin{aligned} \|z_{n+1} - z_n\| &= \|\lambda_{n+1}u + (1 - \lambda_{n+1})T_{n+1}x_{n+1} - (\lambda_n u + (1 - \lambda_n)T_n x_n)\| \\ &\leq |\lambda_{n+1} - \lambda_n|\|u\| + \|T_{n+1}x_{n+1} - T_n x_n\| \\ &\quad + \lambda_n\|T_n x_n\| + \lambda_{n+1}\|T_{n+1}x_{n+1}\| \\ &\leq |\lambda_{n+1} - \lambda_n|\|u\| + \|T_{n+1}x_{n+1} - T_{n+1}x_n\| + \|T_{n+1}x_n - T_n x_n\| \\ &\quad + \|T_n x_n - T_n x_n\| + (\lambda_n + \lambda_{n+1})M \\ &\leq \|x_{n+1} - x_n\| + (|\lambda_{n+1} - \lambda_n| + \lambda_n + \lambda_{n+1})M \\ &\quad + \sup_{x \in C} \|T_{n+1}x - T_n x\| + \sup_{x \in C} \|T_n x - T_n x\|. \end{aligned}$$

Thus, by assumption (B2) and (3.3), we have

$$\begin{aligned} &\limsup_{n \rightarrow \infty} (\|z_{n+1} - z_n\| - \|x_{n+1} - x_n\|) \\ &\leq \lim_{n \rightarrow \infty} \sup_{x \in C} \|T_{n+1}x - T_n x\| + \lim_{n \rightarrow \infty} \sup_{x \in C} \|T_n x - T_n x\| \\ &\quad + \lim_{n \rightarrow \infty} (|\lambda_{n+1} - \lambda_n| + \lambda_n + \lambda_{n+1})M = 0. \end{aligned}$$

By Lemma 2.1, we obtain

$$\lim_{n \rightarrow \infty} \|x_n - z_n\| = 0$$

and hence

$$\lim_{n \rightarrow \infty} \|x_{n+1} - x_n\| = \lim_{n \rightarrow \infty} (1 - \beta_n) \|x_n - z_n\| = 0.$$

Since

$$\begin{aligned} \|x_{n+1} - Tx_n\| &\leq \|x_{n+1} - T_n x_n\| + \|T_n x_n - Tx_n\| \\ &\leq \|x_{n+1} - x_n\| + \|x_n - z_n\| + \|z_n - T_n x_n\| + \sup_{x \in C} \|Tx - T_n x\| \\ &\leq \|x_{n+1} - x_n\| + \|x_n - z_n\| + \lambda_n \|u - T_n x_n\| + \sup_{x \in C} \|Tx - T_n x\|, \end{aligned}$$

then

$$\lim_{n \rightarrow \infty} \|x_{n+1} - Tx_n\| = 0.$$

By Lemma 3.1, there exists  $x^* \in F(T)$  such that

$$\limsup_{n \rightarrow \infty} \langle u - x^*, J(x_{n+1} - x^*) \rangle \leq 0. \quad (3.4)$$

Finally, we show that  $x_n \rightarrow x^* (n \rightarrow \infty)$ . In fact,

$$\begin{aligned} &\|x_{n+1} - x^*\|^2 \\ &= (1 - \alpha_n - \beta_n) \langle T_n x_n - x^*, J(x_{n+1} - x^*) \rangle + \beta_n \langle x_n - x^*, J(x_{n+1} - x^*) \rangle \\ &\quad + \alpha_n \langle u - x^*, J(x_{n+1} - x^*) \rangle \\ &\leq (1 - \alpha_n - \beta_n) \frac{\|T_n x_n - x^*\|^2 + \|J(x_{n+1} - x^*)\|^2}{2} + \beta_n \frac{\|x_n - x^*\|^2 + \|J(x_{n+1} - x^*)\|^2}{2} \\ &\quad + \alpha_n \langle u - x^*, J(x_{n+1} - x^*) \rangle \\ &\leq (1 - \alpha_n) \frac{\|x_n - x^*\|^2}{2} + \frac{\|x_{n+1} - x^*\|^2}{2} + \alpha_n \langle u - x^*, J(x_{n+1} - x^*) \rangle \end{aligned}$$

Therefore,

$$\|x_{n+1} - x^*\|^2 \leq (1 - \alpha_n) \|x_n - x^*\|^2 + 2\alpha_n \langle u - x^*, J(x_{n+1} - x^*) \rangle. \quad (3.5)$$

By the condition (C1), now we apply Lemma 2.5 to yield

$$\lim_{n \rightarrow \infty} \|x_n - x^*\| = 0.$$

The proof is completed.

Using the same proof techniques as Theorem 3.2, the following is obtained easily. Since the proof is a repeating work, we omit it.

**Theorem 3.3** *Let  $E$  be a strictly convex Banach space with a uniformly Gâteaux differentiable norm. Assumed that  $K$  is a nonempty weakly compact convex subset of  $E$  and  $\{T_n\}_{n=1}^{+\infty}$  is a countable family of nonexpansive mappings from  $K$  into itself such that  $F := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$  and the condition (B2). Let  $\{\alpha_n\}$  and  $\{\beta_n\}$  be two real number sequence in  $[0, 1]$  satisfying (C1) and (C3), respectively. Then the modified Mann type iteration sequence  $\{x_n\}$ , defined by (3.1) strongly converges to some point of  $F$ .*

**Theorem 3.4** *Let  $E$  be a reflexive and strictly convex Banach space with a uniformly Gâteaux differentiable norm and  $K$  be a nonempty closed convex subset of  $E$ . Suppose that  $\{T_n\}_{n=1}^{+\infty}$  is a countable family of nonexpansive mappings from  $K$  into itself such that  $F := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$  and the condition (B2). Let  $\{\alpha_n\}$  and*

$\{\beta_n\}$  be two real number sequence in  $[0, 1]$  satisfying (C1) and (C3), respectively. Then the modified Mann type iteration sequence  $\{x_n\}$ , defined by (3.1) strongly converges to some point of  $F$ .

**Proof.** Using the same argumentation as Theorem 3.2, we can show that  $\{x_n\}$  is bounded and (3.2) holds. Since every nonempty closed convex subset is a Chebyshev set in a strictly convex and reflexive Banach space (see [15, Corollary 5.1.19]), then the conclusion of Lemma 3.1 holds also. The desired ultimateness is reached.

**Remark 1** (i) There are many spaces which has the fixed point property for non-expansive self-mappings. For example, uniformly convex Banach space, uniformly smooth Banach space, reflexive Banach space with normal structure, Banach space with Opial's condition and so on.

(ii) We remark that Theorem 3.3 is independent of Theorem 3.2. On the one hand, it is easy to find examples of spaces which satisfies the fixed point property for nonexpansive self-mappings, which are not strictly convex. On the other hand, it appears to be unknown whether a weakly compact convex subset of strictly convex Banach space has the fixed point property for nonexpansive self-mappings.

(iii) In the above theorems, not only the condition (B2) is weaker than (B1), but also the proof is different from ones of Aoyama et al.[1] which isn't dependent upon the convergence of the implicit anchor-like continuous path  $z_t$ , defined by  $z_t = tu + (1 - t)Tz_t$ .

#### 4. Weak convergence of Mann type iteration

Recall that A Banach space  $E$  is said to satisfy *Opial's condition* [16] if for any sequence  $\{x_n\}$  in  $E$ ,  $x_n \rightharpoonup x$  ( $n \rightarrow \infty$ ) implies

$$\limsup_{n \rightarrow \infty} \|x_n - x\| < \limsup_{n \rightarrow \infty} \|x_n - y\|, \forall y \in E \text{ with } x \neq y.$$

Hilbert spaces and  $l^p$  ( $1 < p < \infty$ ) satisfy Opial's condition and Banach spaces with a weakly sequentially continuous duality mapping satisfies Opial's condition [6].

We now show weak convergence of Mann type iteration (1.5) which extend the main result of Aoyama et al.[2] from uniformly convex Banach space to reflexive Banach space.

**Theorem 4.1** *Let  $E$  be a reflexive Banach space satisfying Opial's condition and  $K$  be a nonempty closed convex subset of  $E$ . Suppose  $\{T_n\}$  ( $n = 1, 2, \dots$ ) is a countable family of nonexpansive mappings from  $K$  into itself satisfying the condition (B2) and  $F := \bigcap_{n=1}^{\infty} F(T_n) \neq \emptyset$ . Let  $\{x_n\}$  be a sequence of Mann type iteration defined by (1.5) and  $\alpha_n \in [0, 1]$  satisfy*

$$0 < \liminf_{n \rightarrow \infty} \alpha_n \leq \limsup_{n \rightarrow \infty} \alpha_n < 1.$$

*Then  $\{x_n\}$  weakly converges to some point of  $F$ .*

**Proof.** Take  $p \in F$ . We have

$$\begin{aligned} \|x_{n+1} - p\| &\leq (1 - \alpha_n)\|x_n - p\| + \alpha_n\|T_n x_n - p\| \\ &\leq (1 - \alpha_n)\|x_n - p\| + \alpha_n\|x_n - p\| \\ &\leq \|x_n - p\|. \end{aligned}$$



Then  $\{\|x_n - p\|\}$  is a decreasing sequence, and hence  $\lim_{n \rightarrow \infty} \|x_n - p\|$  exists for each  $p \in F$  and  $\{x_n\}$  is bounded. Let  $C$  be any bounded subset of  $K$  containing  $\{x_n\}$ , then

$$\begin{aligned} \|T_{n+1}x_{n+1} - T_nx_n\| &\leq \|T_{n+1}x_{n+1} - T_{n+1}x_n\| + \|T_{n+1}x_n - Tx_n\| \\ &\quad + \|Tx_n - T_nx_n\| \\ &\leq \|x_{n+1} - x_n\| + \sup_{x \in C} \|T_{n+1}x - Tx\| \\ &\quad + \sup_{x \in C} \|Tx - T_nx\|. \end{aligned}$$

It follows from the hypothesis that

$$\begin{aligned} &\limsup_{n \rightarrow \infty} (\|T_{n+1}x_{n+1} - T_nx_n\| - \|x_{n+1} - x_n\|) \\ &\leq \limsup_{n \rightarrow \infty} \sup_{x \in C} \|T_{n+1}x - Tx\| + \limsup_{n \rightarrow \infty} \sup_{x \in C} \|Tx - T_nx\| = 0. \end{aligned}$$

By Lemma 2.1, we obtain

$$\lim_{n \rightarrow \infty} \|x_n - T_nx_n\| = 0.$$

Thus, we have

$$\begin{aligned} \|x_n - Tx_n\| &\leq \|x_n - T_nx_n\| + \|T_nx_n - Tx_n\| \\ &\leq \|x_n - T_nx_n\| + \limsup_{n \rightarrow \infty} \sup_{x \in C} \|Tx - T_nx\| = 0. \end{aligned}$$

Hence,

$$\lim_{n \rightarrow \infty} \|x_n - Tx_n\| = 0.$$

Since  $E$  is reflexive, there exists a subsequence  $\{x_{n_k}\}$  of  $\{x_n\}$  such that  $x_{n_k} \rightharpoonup x^*$  for some  $x^* \in K$ . Then  $x^* \in F$ . In fact, suppose not. Then the Opial's property of  $E$  implies the following:

$$\begin{aligned} \limsup_{k \rightarrow \infty} \|x_{n_k} - x^*\| &< \limsup_{k \rightarrow \infty} \|x_{n_k} - Tx^*\| \\ &\leq \limsup_{k \rightarrow \infty} (\|x_{n_k} - T_nx_k\| + \|Tx_{n_k} - Tx^*\|) \\ &= \limsup_{k \rightarrow \infty} \|x_{n_k} - x^*\|. \end{aligned}$$

This gets a contradiction. Hence  $x^* = Tx^* \in F$ .

Next we show  $x_n \rightharpoonup x^*$ . Suppose not. There exists another subsequence  $\{x_{n_i}\}$  of  $\{x_n\}$  such that  $x_{n_i} \rightharpoonup x \neq x^*$ . Then, we also have  $x = Tx$ . From Opial's property, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \|x_n - x\| &= \limsup_{i \rightarrow \infty} \|x_{n_i} - x\| \\ &< \limsup_{i \rightarrow \infty} \|x_{n_i} - x^*\| = \limsup_{k \rightarrow \infty} \|x_{n_k} - x^*\| \\ &< \limsup_{k \rightarrow \infty} \|x_{n_k} - x\| = \lim_{n \rightarrow \infty} \|x_n - x\|. \end{aligned}$$

Which gets a contradiction. So the conclusion of the theorem follows.

**Remark 2.** It isn't known whether the assumption (B2) can be displaced by the weaker condition (B3).

(B3) For any bounded subset  $C$  of  $K$ , there exists a nonexpansive mapping  $T$  of  $K$  into itself and a subsequence  $\{T_{n_i}\}$  of  $\{T_n\}$  such that

$$\lim_{i \rightarrow \infty} \sup_{x \in C} \|Tx - T_{n_i}x\| = 0 \text{ and } F(T) = F.$$

## REFERENCES

1. K. Aoyama, Y. Kimura, W. Takahashi and M. Toyoda, *Approximation of common fixed points of a countable family of nonexpansive mappings in a Banach space*, Nonlinear Analysis, 67(2007) 2350–2360.
2. K. Aoyama, Y. Kimura, W. Takahashi and M. Toyoda, *Finding common fixed points of a countable family of nonexpansive mappings in a Banach space*, Scientiae Mathematicae Japonicae, 66(2007) 89–99 :e2007, 325–335.
3. F.E. Browder, *Fixed point theorems for noncompact mappings in Hilbert space*, Proc. Nat. Acad. Sci. U.S.A. 53 (1965) 1272–1276.
4. F.E. Browder, *Nonlinear operators and nonlinear equations of evolution in Banach spaces*, Nonlinear Functional Analysis (Proc. Sympos. PureMath., Vol. 18, Part 2, Chicago, Ill., 1968), American Mathematical Society, Rhode Island, 1976, pp. 1–308.
5. C. E. Chidume and C. O. Chidume, *Iterative approximation of fixed points of nonexpansive mappings*, J. Math. Anal Appl. 318(2006) 288–295.
6. J.P. Gossez and E.L. Dozo, *Some geometric properties related to the fixed point theory for nonexpansive mappings*, Pacific J. Math. 40(1972), 565–573.
7. B. Halpern, *Fixed points of nonexpansive maps*, Bull. Amer. Math. Soc. 73(1967) 957–961.
8. V.I. Istratescu, *Fixed Point Theory: An Introduction*, Published by D.Reidel Publishing Company (1981), The Netherlands.
9. J.S. Jung, *Iterative approaches to common fixed points of nonexpansive mappings in Banach spaces*, J. Math. Anal Appl. 302(2005) 509–520.
10. T.H. Kim and H.K. Xu, *Strong convergence of modified Mann iterations*, Nonlinear Anal. 61(2005) 51–60.
11. W. A. Kirk and S. Massa, *Remarks on asymptotic and Chebyshev centers*, Houston J. Math. 16 (1990), no. 3, 357–364.
12. W. A. Kirk, *Transfinite methods in metric fixed point theory*. Abstract and Applied Analysis, 2003:5(2003) 311–324.
13. T. C. Lim, *Remarks on some fixed point theorems*, Proc. Amer. Math. Soc. 60 (1976), 179–182.
14. W.R. Mann, *Mean value methods in iteration*, Proc. Amer. Math. Soc. 4(1953) 506–510.
15. R. E. Megginson, *An introduction to Banach space theory*, 1998 Springer-Verlag New York, Inc.
16. Z. Opial, *Weak convergence of the sequence of successive approximations for nonexpansive mappings*, Bull. Amer. Math. Soc. 73 (1967), 591–597.
17. J. G. O'Hara, P. Pillay and H.K. Xu, *Iterative approaches to finding nearest common fixed point of nonexpansive mappings in Hilbert spaces*, Nonlinear Analysis, 54(2003) 1417–1426.
18. J. G. O'Hara, P. Pillay and H.K. Xu, *Iterative Approaches to Convex Feasibility Problem in Banach Space*, Nonlinear Analysis, 64(2006) 2022–2042.
19. S. Reich, *Strong convergence theorems for resolvents of accretive operators in Banach spaces*, J. Math. Anal Appl. 75(1980) 287–292.
20. Y. Song, *A new sufficient condition for the strong convergence of Halpern type iterations*, Applied Mathematics and Computation, 198(2)(2008), 721–728.
21. Y. Song, *Yisheng Song, On a Mann type implicit iteration process for continuous pseudo-contractive mappings*, Nonlinear Analysis, 67(2007) 3058–3063.
22. Y. Song, *Iterative selection methods for the common fixed point problems in a Banach space*, Applied Mathematics and Computation, 193(2007) 7–17.
23. Y. Song, *Weak and strong convergence of Mann's-type iterations for a countable family of nonexpansive mappings*, Journal of the Korean Mathematical Society (Accepted).
24. Y. Song, *Iterative approximation of a countable family of nonexpansive mappings*, Applicable Analysis, 86(2007), 1329–1337.
25. Y. Song and R. Chen, *Strong convergence of an iterative method for non-expansive mappings*, Mathematische Nachrichten, 281(8)(2008), 1–9.
26. Y. Song and Y.J. Cho, *Iterative approximations for multivalued nonexpansive mappings in reflexive Banach spaces*, Math. Inequal. Appl.(Accepted).
27. Y. Song and R. Chen, *Strong convergence theorems on an iterative method for a family of finite nonexpansive mappings*, Applied Mathematics and Computation, 180 (2006) 275–287.
28. Y. Song and R. Chen, *Viscosity approximation methods for nonexpansive nonself-mappings*, J. Math. Anal. Appl. 321(2006), 316–326.

## COUNTABLE FAMILY OF NONEXPANSIVE MAPPINGS

29. Y. Song and R. Chen, *Iterative approximation to common fixed points of nonexpansive mapping sequences in reflexive Banach spaces*, Nonlinear Analysis, 66 (2007) 591–603.
30. Y. Song, R. Chen and H. Zhou, *Viscosity approximation methods for nonexpansive mapping sequences in Banach spaces*, Nonlinear Analysis, 66(2007) 1016–1024.
31. Y. Song and S. Xu, *Strong convergence theorems for nonexpansive semigroup in Banach spaces*, J. Math. Anal. Appl., 338(2008) 152–161.
32. N. Shioji and W. Takahashi, *Strong convergence of approximated sequences for nonexpansive mappings in Banach spaces*, Proc. Amer. Math. Soc., 125(1997) 3641–3645.
33. T. Suzuki, *Strong convergence theorems for infinite families of nonexpansive mappings in general Banach spaces*, Fixed Point Theory and Applications, 2005(1)(2005) 103–123. doi:10.1155/FPTA.2005.103
34. W. Takahashi, *Nonlinear Functional Analysis– Fixed Point Theory and its Applications*, Yokohama Publishers inc, Yokohama, 2000(English).
35. W. Takahashi and Y. Ueda, *On Reich's strong convergence for resolvents of accretive operators*, J. Math. Anal. Appl. 104(1984) 546–553.
36. H.K. Xu, *Iterative algorithms for nonlinear operators*, J. London Math. Soc. 66 (2002) 240–256.

STATISTICAL CONVERGENCE AND STATISTICAL CORE OF SEQUENCES OF  
BOUNDED LINEAR OPERATORS

A. GÖKHAN

Fırat University, Faculty of Science, Department of Mathematics, 23119,  
Elazığ, Turkey.

E-mail:agokhan1@firat.edu.tr.

**Abstract:** In this study, we introduce a notion of uniform, strong and weak statistical convergence of sequences of bounded linear operators. We also give the relations between these convergences and uniform operator convergence, strong operator convergence, weak operator convergence. Furthermore we introduce the concept of the statistical Cauchy sequence for sequences of bounded linear operators and prove that it is equivalent to statistical convergence of sequences of bounded linear operators. In addition to these results we study statistical core for bounded linear operators and give some inequalities.

**Key Words:** Operator sequence; Uniform operator convergence; Strong operator convergence; Weak operator convergence; Statistical convergence; Pointwise statistical convergence, Core theorems and Matrix Transformations.

**MSC 2000:** 40A05, 40C05

## 1. INTRODUCTION

The idea of statistical convergence was introduced by Fast [2] and Schoenberg [7] independently. Later on it was studied by Fridy [3], Salat [6], and Tripathy [8] and many others. Gökhan and Güngör [5] defined pointwise statistical convergence of sequences of real-valued functions. Connor [1] gave an extension of the notion of statistical convergence where the asymptotic density is replaced by a finitely additive set function  $\mu$ .

Let  $\mathbf{N}$  be the space of natural numbers. For each  $E \subset \mathbf{N}$ , let  $K_n(E)$  be the cardinality of the set  $E \cap [0, n]$ . The asymptotic (or natural) density of  $E$  is given by  $\delta(E) = \lim_{n \rightarrow \infty} \frac{K_n(E)}{n}$  whenever the limit exists. Clearly finite sets have zero density,  $\delta(E^c) = \delta(\mathbf{N} - E) = 1 - \delta(E)$ , whenever both sides exist, where  $E^c$  is the complement of the set  $E$  in  $\mathbf{N}$ . We say that a real number sequence  $(x_n)$  is statistical convergent to  $\ell$  provided that for every  $\varepsilon > 0$ ,  $\delta(\{k \in \mathbf{N} : |x_k - \ell| \geq \varepsilon\}) = 0$  or,  $\delta(\{k \in \mathbf{N} : |x_k - \ell| < \varepsilon\}) = 1$  for every  $\varepsilon > 0$  in which case we write  $\text{st-lim } x_k = \ell$ .

Let  $X$  and  $Y$  be normed spaces and  $B(X, Y)$  be the normed spaces of all bounded linear operators from  $X$  into  $Y$  with the usual operator norm.

As we know, a sequence of operators  $T_k \in B(X, Y)$  tends to a limit  $T$ , where  $T : X \rightarrow Y$  is an operator, if given  $\varepsilon > 0$ , we can find an integer  $k_0$  such that

- i)  $\|T_k - T\| < \varepsilon$ , for all  $k > k_0$ ,
- ii)  $\|T_k x - T x\| < \varepsilon$ , for all  $k > k_0$  and for every  $x \in X$ ,

iii)  $\|f(T_k x) - f(Tx)\| < \varepsilon$ , for all  $k > k_0$ , for every  $x \in X$  and  $f \in Y'$ , where  $Y'$  denotes the set of all bounded linear functional on  $Y$ . Then  $T$  is called the uniform operator limit, strong operator limit and weak operator limit of  $(T_k)$ , respectively. It is well known that  $(i) \implies (ii) \implies (iii)$ .

## 2. STATISTICAL CONVERGENCE OF SEQUENCES OF OPERATORS

Some operator sequences does not convergence in above convergence modes but its might converge in a weaker sense. Therefore, in the present paper, we introduce a notion of uniform, strong and weak statistical convergence of sequences of bounded linear operators.

Throughout the paper,  $(T_k)$  will denote a sequence of operators  $T_k \in B(X, Y)$ .

**Definition 2.1.** The sequence  $(T_k)$  is said to be uniformly statistical operator convergent to  $T$ , if for every  $\varepsilon > 0$ , there exists such a set  $E \subset \mathbf{N}$  that  $\delta(E) = 1$  and  $\exists k_0(\varepsilon) \in E \ni \forall k > k_0$  and  $k \in E$ ,  $\|T_k - T\| < \varepsilon$ ,  
i.e. for every  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} |\{k \leq n : \|T_k - T\| \geq \varepsilon\}| = 0.$$

In this case we write  $\text{st-lim} T_k = T$  or  $T_k \xrightarrow{\text{st.}} T$ , where  $T$  is an operator from  $X$  into  $Y$ .

**Definition 2.2.** The sequence  $(T_k)$  is said to be strongly statistical operator convergent to  $T$ , if for  $\forall \varepsilon > 0$  and for  $\forall x \in X$ , there exists such a set  $E_x \subset \mathbf{N}$  that  $\delta(E_x) = 1$  and  $\exists k_0 \in E_x \ni \forall k > k_0$  and  $k \in E_x$ ,  $\|T_k x - Tx\| < \varepsilon$ ,  
i.e. for every  $\varepsilon > 0$  and for every  $x \in X$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} |\{k \leq n : \|T_k x - Tx\| \geq \varepsilon\}| = 0.$$

In this case, we write  $\text{st-lim} T_k x = Tx$  or  $T_k x \xrightarrow{\text{st.}} Tx$  on  $Y$  for every  $x \in X$ , where  $T$  is an operator from  $X$  into  $Y$ .

**Definition 2.3.** The sequence  $(T_k)$  is said to be weakly statistical operator convergent to  $T$ , if  $\forall \varepsilon > 0$ ,  $\forall x \in X$  and  $\forall f \in Y'$ , there exists such a set  $E_{x,f} \subset \mathbf{N}$  that  $\delta(E_{x,f}) = 1$  and  $\exists k_0 \in E_{x,f} \ni \forall k > k_0$  and  $k \in E_{x,f}$ ,  
 $|f(T_k x) - f(Tx)| < \varepsilon$ ,

i.e. for every  $\varepsilon > 0$  and for every  $x \in X$  and every  $f \in Y'$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} |\{k \leq n : |f(T_k x) - f(Tx)| \geq \varepsilon\}| = 0$$

In this case, we write  $\text{w.st-lim} T_k x = Tx$  or  $T_k x \xrightarrow{\text{w.st.}} Tx$  on  $Y$  for every  $x \in X$ , where  $T$  is an operator from  $X$  into  $Y$ .

Since the proof of the following theorems is obvious, we merely state its and omit its.

**Theorem 2.1:** Let  $(T_k)$  and  $(S_k)$  be two sequences of operators.

- i) If  $T_k x \xrightarrow{\mu} Tx$  and  $S_k x \xrightarrow{\mu} Sx$  on  $Y$  for every  $x \in X$ , then  $\alpha T_k x + \beta S_k x \xrightarrow{\mu} \alpha Tx + \beta Sx$  on  $Y$  for every  $x \in X$ , where  $\alpha, \beta \in \mathbb{R}$ .
- ii) If  $T_k x \xrightarrow{\mu} Tx$  on  $Y$  for every  $x \in X$ , then  $\|T_k x\| \xrightarrow{\mu} \|Tx\|$  for every  $x \in X$ ,  
where  $\mu = \text{st.}$  or  $\text{w.st.}$

**Theorem 2.2:** Let  $(T_k)$  and  $(S_k)$  be two sequences of operators.

- i) If  $T_k \xrightarrow{\text{st.}} T$  and  $S_k \xrightarrow{\text{st.}} S$ , then  $\alpha T_k + \beta S_k \xrightarrow{\text{st.}} \alpha T + \beta S$ , where  $\alpha, \beta \in \mathbb{R}$ .
- ii) If  $T_k \xrightarrow{\text{st.}} T$ , then  $\|T_k\| \xrightarrow{\text{st.}} \|T\|$ .

### 3. RELATIONS BETWEEN MODES OF CONVERGENCE

It is not difficult to show that

- i)  $T_k \rightarrow T \Rightarrow T_k \xrightarrow{\text{st.}} T$ ,
- ii)  $T_k x \rightarrow Tx$  on  $Y$  for every  $x \in X \Rightarrow T_k x \xrightarrow{\text{st.}} Tx$  on  $Y$  for every  $x \in X$ ,
- iii)  $T_k x \xrightarrow{w} Tx$  on  $Y$  for every  $x \in X \Rightarrow T_k x \xrightarrow{w, \text{st.}} Tx$  on  $Y$  for every  $x \in X$ .

But we note that the converses of (i), (ii) and (iii) are not true. The following examples are provided to clarify these.

**Example 3.1.** Now, let us consider a sequence  $(T_k)$  of operators  $T_k : \ell_\infty \rightarrow c$  is defined by

$$T_k x = \begin{cases} (\xi_1, \frac{\xi_2}{2}, \frac{\xi_3}{3}, \dots), & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ (0, 0, 0, \dots), & \text{otherwise} \end{cases}$$

where,  $x = (\xi_1, \xi_2, \xi_3, \dots) \in \ell_\infty$ . The operator  $T_k$  is linear and bounded for every  $k \in \mathbb{N}$ .  $(T_k)$  is uniformly statistical operator convergent to  $T = 0$  since

$\frac{1}{n} |\{k \leq n : \|T_k - 0\| \geq \varepsilon\}| \leq \frac{p(p+1)}{2 \cdot 3^p}$   
for every  $p, n \in \mathbb{N}$  such that  $3^p \leq n < 3^{p+1}$ , where

$$\|T_k - 0\| = \begin{cases} 1, & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

for each  $k \in \mathbb{N}$ . However,  $(T_k)$  is not uniformly operator convergent to  $T = 0$  since  $\lim_{k \rightarrow \infty} \|T_k\|$  does not exist.

**Example 3.2.** A sequence  $(T_k)$  of operators  $T_k : \ell^2 \rightarrow \ell^2$  is defined by

$$T_k x = \begin{cases} \underbrace{(0, 0, \dots, 0)}_{(k \text{ zeros})}, \xi_1, \xi_2, \xi_3, \dots, & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ (0, 0, 0, \dots), & \text{otherwise} \end{cases}$$

where,  $x = (\xi_1, \xi_2, \xi_3, \dots) \in \ell^2$ . This operator  $T_k$  is linear and bounded for every  $k \in \mathbb{N}$ . We show that  $(T_k)$  is strongly statistical operator convergent to  $T = 0$  since

$$\frac{1}{n} |\{k \leq n : \|T_k x - 0x\| \geq \varepsilon\}| \leq \frac{p(p+1)}{2 \cdot 3^p}$$
 for every  $p, n \in \mathbb{N}$  such that  $3^p \leq n < 3^{p+1}$  and for every  $x \in \ell^2$ . So  $\text{st-lim } T_k x = 0$  on  $\ell^2$ . However  $(T_k)$  is not strongly operator convergent because for  $x = (1, 0, 0, \dots) \in \ell^2$  we have

$$\|T_m x - T_n x\| = \begin{cases} \sqrt{2}, & m \neq n \text{ and } n \in I_p \text{ for some } p \in \mathbb{N}, m \in I_r \text{ for some } r \in \mathbb{N} \\ 1, & n \in I_p \text{ and } m \notin I_p \text{ for some } p \in \mathbb{N} \\ 0, & m, n \notin I_p \text{ for all } p \in \mathbb{N} \text{ or } m=n \end{cases}$$

where  $I_p = [3^p, 3^p + p)$ .

**Example 3.3.** In the space  $\ell^2$ , we consider a sequence  $(T_k)$ , where  $T_k : \ell^2 \rightarrow \ell^2$  is defined by

$$T_k x = \begin{cases} (\xi_1, \xi_2, \dots), & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ (0, 0, 0, \dots), & \text{otherwise} \end{cases}$$

where,  $x = (\xi_1, \xi_2, \xi_3, \dots) \in \ell^2$  and  $T_k \in B(X, Y)$ . It is easy to see that  $(T_k)$  is weakly statistical operator convergent to 0 since

$$f(T_k x) = \langle T_k x, z \rangle = \begin{cases} \sum_{i=1}^{\infty} \xi_i \bar{\zeta}_i, & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ 0, & \text{otherwise} \end{cases}$$

by Riesz representation, where  $z = (\zeta_i) \in \ell^2$ . However it is easy to see that  $f(T_k x) \not\rightarrow f(0x)$ .

We now proceed to examine some relationships between uniformly statistical operator convergence, strongly statistical operator convergence and weakly statistical operator convergence.

**Theorem 3.1:** Uniform statistical operator convergence implies strong statistical operator convergence with the same limit. But the converse of this theorem is not true, as the following example shows:

**Example 3.4.** In the space  $\ell^1$ , we consider a sequence  $(T_k)$ , where  $T_k : \ell^1 \rightarrow \ell^1$  is defined by

$$T_k x = \begin{cases} (\xi_1, \xi_2, \dots), & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ (0, 0, \dots, 0, 2\xi_k, 2\xi_{k+1}, 2\xi_{k+2}, \dots), & \text{otherwise} \end{cases}$$

where,  $x = (\xi_1, \xi_2, \xi_3, \dots) \in \ell^1$ . This operator  $T_k$  is linear and bounded for every  $k \in \mathbb{N}$ . Clearly, for every  $x \in \ell^1$ ,

$\frac{1}{n}|\{k \leq n : \|T_k x - 0x\| \geq \varepsilon\}| \rightarrow 0$ ,  
i.e.  $\text{st-lim} T_k x = 0x$  on  $\ell^1$ . But  $(T_k)$  is not uniformly statistical operator convergent to  $T = 0$  since

$$\|T_k - 0\| = \begin{cases} 1, & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ 2, & \text{otherwise} \end{cases}$$

**Theorem 3.2.** Strong statistical operator convergence implies weak statistical operator convergence with the same limit.

We note that the converse is not true. For this let us consider the following example.

**Example 3.5.** A sequence  $(T_k)$  of operators  $T_k : \ell^2 \rightarrow \ell^2$  is defined by

$$T_k x = \begin{cases} (\xi_1, \xi_2, \xi_3, \dots), & k \in [3^p, 3^p + p), p = 1, 2, \dots \\ (\underbrace{0, 0, \dots, 0}_{(k \text{ zeros})}, \xi_1, \xi_2, \xi_3, \dots), & \text{otherwise} \end{cases}$$

where  $x = (\xi_1, \xi_2, \xi_3, \dots) \in \ell^2$ . This operator  $T_k$  is linear and bounded. We show that  $(T_k)$  is weakly statistical operator convergent to 0 but not strongly statistical operator convergent.

Every bounded linear functional  $f$  on  $\ell^2$  has a Riesz representation. Hence, setting  $i = k + j$  and using the definition of  $T_k$ , we have

$$f(T_k x) = \langle T_k x, z \rangle = \begin{cases} \sum_{i=1}^{\infty} \xi_i \bar{\zeta}_i, & k \in [3^p, 3^p + p), p = 1, 2, 3, \dots \\ \sum_{i=k+1}^{\infty} \xi_{i-k} \bar{\zeta}_i = \sum_{j=1}^{\infty} \xi_j \bar{\zeta}_{k+j}, & \text{otherwise} \end{cases}$$

where  $z = (\zeta_i) \in \ell^2$ . By the Cauchy-Schwarz inequality

$$|f(T_k x)|^2 = |\langle T_k x, z \rangle|^2 \leq \begin{cases} \sum_{i=1}^{\infty} |\xi_i|^2 \sum_{i=1}^{\infty} |\zeta_i|^2, & k \in [3^p, 3^p + p), p = 1, 2, 3, \dots \\ \sum_{j=1}^{\infty} |\xi_j|^2 \sum_{m=k+1}^{\infty} |\zeta_m|^2, & \text{otherwise} \end{cases}$$

Since the set  $[3^p, 3^p + p)$  has density zero, it follows that  $f(T_k x) \xrightarrow{st} 0 = f(0x)$  by Theorem 2 in [8]. Consequently,  $(T_k)$  is weakly statistical operator convergent to 0. However  $(T_k)$  is not strongly statistical operator convergent to 0, because we have

$$\|T_k x - 0x\| = \begin{cases} (\sum_{i=1}^{\infty} |\xi_i|^2)^{1/2}, & k \in [3^p, 3^p + p), p = 1, 2, 3, \dots \\ (\sum_{i=k+1}^{\infty} |\xi_{i-k}|^2)^{1/2}, & \text{otherwise} \end{cases}$$

and so  $\|T_k x - 0x\| \not\xrightarrow{st} 0$ .



We know that in the finite dimensional normed spaces, the strong and weak convergence are equivalent concepts. The below result is a statistical analogue of this result.

Suppose that  $T_k x \xrightarrow{w.st.} Tx$  and  $\dim Y = n$ . Let  $\{e_1, e_2, \dots, e_n\}$  be any basis for  $Y$  and  $\{f_1, f_2, \dots, f_n\}$  be its dual basis and, say  $T_k x = \alpha_1^{(k)} e_1 + \dots + \alpha_n^{(k)} e_n$  and  $Tx = \alpha_1 e_1 + \dots + \alpha_n e_n$  for every  $x \in X$ . Then  $f_j(T_k x) = \alpha_j^{(k)}$ ,  $f_j(Tx) = \alpha_j$ . Hence and by our assumption  $f_j(T_k x) \xrightarrow{st.} f_j(Tx)$  for fixed  $f_j (j = 1, \dots, n)$  implies  $\alpha_j^{(k)} \xrightarrow{st.} \alpha_j$  for  $j = 1, \dots, n$ . From this we readily obtain

$$\|T_k x - Tx\| \leq \sum_{j=1}^n \left| \alpha_j^{(k)} - \alpha_j \right| \cdot \|e_j\| \xrightarrow{st.} 0$$

for every  $x \in X$ . This shows that  $(T_k)$  converges strongly statistically to  $T$ .

#### 4. STATISTICAL CAUCHY SEQUENCES OF OPERATORS

Now, we introduce the statistical analog of the Cauchy convergence criterion for the sequences of bounded linear operators.

**Definition 4.1.** The sequence  $(T_k)$  is a uniformly statistical operator Cauchy sequence provided that for every  $\varepsilon > 0$ , there exists such a set  $E \subset \mathbf{N}$  that  $\delta(E) = 1$  and  $\exists k_0(\varepsilon) \in E$  and an  $N(= N(\varepsilon)) \ni \forall k > k_0, k \in E, \|T_k - T_N\| < \varepsilon$ .

**Definition 4.2.** The sequence  $(T_k)$  is a strongly statistical operator Cauchy sequence provided that for every  $\varepsilon > 0$ , for every  $x \in X$  there exists such a set  $E_x \subset \mathbf{N}$  that  $\delta(E_x) = 1$  and  $\exists k_0(= k_0(\varepsilon, x)) \in E_x$  and an  $N(= N(\varepsilon, x)) \ni \forall k > k_0, k \in E_x, \|T_k x - T_N x\| < \varepsilon$ .

**Definition 4.3.** The sequence  $(T_k)$  is a weakly statistical operator Cauchy sequence provided that  $\forall \varepsilon > 0, \forall x \in X$  and  $\forall f \in Y^*$ , there exists such a set  $E_{x,f} \subset \mathbf{N}$  that  $\delta(E_{x,f}) = 1$  and  $\exists k_0 \in E_{x,f}$  and an  $N(= N(\varepsilon)) \ni \forall k > k_0$  and  $k \in E_{x,f}, |f(T_k x) - f(T_N x)| < \varepsilon$ .

**Theorem 4.1.** Let  $Y$  be a Banach space. Then  $(T_k)$  is strongly statistical operator convergent on  $X$  if and only if  $(T_k)$  is a strongly statistical Cauchy sequence on  $X$ .

**Proof:** Assume that  $\text{st-lim } T_k x = Tx$  for every  $x \in X$ . Then for every  $\varepsilon > 0$  and for every  $x \in X$  there exists such a set  $E \subset \mathbf{N}$  that  $\delta(E_x) = 1$  and  $\exists k_0 \in E_x \ni \forall k \geq k_0, k \in E_x, \|T_k x - Tx\| < \varepsilon/2$ . If  $N$  is chosen so that  $\|T_N x - Tx\| < \varepsilon/2$  for every  $x \in X$  then we have  $\|T_k x - T_N x\| < \varepsilon$  for  $\exists k_0 \in E_x, \forall k > k_0, k \in E_x$  and for every  $x \in X$ .

Now suppose that  $(T_k)$  is strongly statistical Cauchy sequence. Then we may choose  $n(1)$  such that the set  $A_1 = \{k : \|T_k x - T_{n(1)} x\| < 1\}$  has asymptotic density 1 for every  $x \in X$ . Suppose that  $n(1) < n(2) < \dots < n(p)$  have been selected, if  $1 \leq r \leq s \leq p$  and  $A_s = \{k : \|T_k x - T_{n(s)} x\| < 2^{1-s}\}$ , then the set  $A_r$

has asymptotic density 1 for every  $x \in X$  and  $n(s) \in A_r$ . Choose  $N$  such that the set  $\{k : \|T_k x - T_N x\| < 2^{-(p+1)}\}$  has asymptotic density 1 for every  $x \in X$ .

Since the set  $(\bigcap_{j=1}^N A_j) \cap \{k : \|T_k x - T_N x\| < 2^{-(p+1)}\}$  has asymptotic density

1, there exists an  $n(p+1) \in (\bigcap_{j=1}^N A_j) \cap \{k : \|T_k x - T_N x\| < 2^{-(p+1)}\}$  such that

$n(p) < n(p+1)$  and  $A_{p+1} = \{k : \|T_k x - T_{n(p+1)} x\| < 2^{-p}\} \supseteq \{k : \|T_k x - T_N x\| < 2^{-(p+1)}\}$  for every  $x \in X$ . Observe that  $A_{p+1}$  has asymptotic density 1 and  $n(p+1) \in A_s$  for all  $s \leq p+1$  and for every  $x \in X$ .

Note that since  $\|T_{n(p)} x - T_{n(p+1)} x\| < 2^{-p}$  for every  $x \in X$ ,  $(T_{n(p)})$  is Cauchy sequence on  $Y$ . Since  $Y$  is Banach space, there exists a  $Tx \in Y$  such that  $\lim_{p \rightarrow \infty} T_{n(p)} x = Tx$ . We claim that  $(T_k)$  is strongly statistical convergent to  $Tx$  for every  $x \in X$ . Let  $\varepsilon > 0$  be given and select  $p \in \mathbb{N}$  such that  $\|T_{n(p)} x - Tx\| < \frac{\varepsilon}{2}$  for every  $x \in X$  and  $\varepsilon > 2^{-p}$ . Note that if  $\|T_k x - Tx\| \geq \varepsilon$  then  $\|T_{n(p)} x - T_k x\| > \frac{\varepsilon}{2} > 2^{1-p}$ , and hence  $k$  is not element of  $A_p$ . It follows that  $\{k : \|T_k x - Tx\| \geq \varepsilon\}$  has asymptotic density zero for every  $x \in X$ . Hence  $(T_k)$  is strongly statistical operator convergent on  $X$ .

**Theorem 4.2.** Let  $Y$  be a Banach space. Then  $(T_k)$  is uniformly statistical operator convergent on  $X$  if and only if  $(T_k)$  is a uniformly statistical Cauchy sequence on  $X$ .

**Proof:** It can be shown in a similar way of Theorem 4.1. Therefore, we omit it.

The next theorems are statistical analogues of a well-known theorems

**Theorem 4.3.** Let every statistical Cauchy sequence in  $Y$  be statistical convergent sequence. Then every statistical Cauchy sequence in  $B(X, Y)$  is statistical convergent.

**Proof:** We consider an arbitrary statistical Cauchy sequence  $(T_k)$  in  $B(X, Y)$  and show that  $(T_k)$  statistical converges to an operator  $T \in B(X, Y)$ . Since  $(T_k)$  is statistical Cauchy, for every  $\varepsilon > 0$  and there exists such a set  $E \subset \mathbb{N}$  that  $\delta(E) = 1$  and  $\exists k_0 \in E$ ,  $\exists N \in \mathbb{N} \forall k > k_0, k \in E$ ,  $\|T_k - T_N\| < \varepsilon$ . For all  $x \in X$  and  $\forall k > k_0, k, k_0 \in E$ , we thus obtain

$$\|T_k x - T_N x\| < \varepsilon \|x\| \quad (1)$$

Now for any fixed  $x$  and given  $\varepsilon_1$ , we may choose  $\varepsilon = \varepsilon_x$  so that  $\varepsilon_x \|x\| < \varepsilon$ . From (1), we see that  $(T_k x)$  is a statistical Cauchy sequence in  $Y$ . Then  $(T_k x)$  is statistical convergent, say,  $T_k x \xrightarrow{st} y$ . Clearly, the statistical limit  $y \in Y$  depends on the choice of  $x \in X$ . This defines an operator  $T : X \rightarrow Y$ , where  $y = Tx$ . Since

$T(\alpha x + \beta y) = st\text{-}\lim T_k(\alpha x + \beta y) = \alpha st\text{-}\lim T_k x + \beta st\text{-}\lim T_k y = \alpha Tx + \beta Ty$ , the operator  $T_k$  is linear. Using the continuity of the norm and Lemma 5 [7], we obtain from (1) for every  $\varepsilon > 0$  and  $\forall k > k_0, k \in E$ , an  $N \in \mathbb{N}$  and all  $x \in X$ ,

$$\begin{aligned}
\|Tx\| &= \|T_Nx - (T_Nx - Tx)\| \\
&\leq \|T_Nx\| + \|T_Nx - \text{st-lim } T_kx\| \\
&\leq c\|x\| + \text{st-lim } \|T_Nx - T_kx\| \leq (c + \varepsilon)\|x\|
\end{aligned}$$

since  $T_N$  is bounded. This shows that  $T$  is a bounded linear operator. Furthermore, we obtain  $\|T_k - T\| = \sup_{\|x\|=1} \|T_kx - Tx\| \xrightarrow{\text{st}} 0$  from  $T_kx \xrightarrow{\text{st}} Tx$  for every  $x \in X$ .

**Definition 4.4.** The sequence  $(T_k)$  is said to be statistical bounded if there exists such a set  $E \subset \mathbf{N}$  that  $\delta(E) = 1$  and exists a  $M_x > 0$ , such that  $\|T_kx\| \leq M_x$  for all  $k \in E$  and every  $x \in X$ . If  $M_x = M$ , then  $(T_k)$  is said to be uniformly statistical bounded.

**Lemma 4.1:** Let  $(T_k)$  be a sequence of operators  $T_k \in B(X, Y)$  and  $\overline{B}_r(x_0) \subset X$  be a closed ball such that  $(\|T_kx\|)$  is statistical uniform bounded for  $\forall x \in \overline{B}_0 = \overline{B}_r(x_0)$ , say,

$$\sup_{k \in E} \|T_kx\| = c,$$

where  $c$  is a real number and  $\delta(E) = 1$ . Then the sequence  $(\|T_k\|)$  is statistical bounded.

**Proof:** Let  $x \in X$  be arbitrary, not zero. Then  $\left\|x_0 + \frac{rx}{\|x\|} - x_0\right\| < r$ , so that  $x_0 + \frac{rx}{\|x\|} \in B_0$ . This yields for all  $k \in E$ , where the set  $E \subset \mathbf{N}$  and  $\delta(E) = 1$ ,  $\|T_kx\| \leq \frac{2c}{r}\|x\|$ . Hence for all  $k \in E$ ,  $\|T_k\| = \sup_{\|x\|=1} \|T_kx\| \leq \frac{2c}{r}$ .

**Theorem 4.4.** Let  $(T_k) \in B(X, Y)$ , where  $X$  is a Banach space and  $Y$  a normed space. If there exists a set  $E \subset \mathbf{N}$  of asymptotic density 1 such that for  $\forall x \in X$  and for  $\forall k \in E$ ,

$$\|T_kx\| \leq c_x$$

where  $c_x$  is a real number, then there is a  $c$  such that

$$\|T_k\| \leq c$$

for all  $k \in E$ .

**Proof:** We assume that the sequence of the norms  $\|T_k\|$  is not bounded on the set  $E$ . Then the sequence  $(\|T_kx\|)$  is not also bounded on the set  $E$  and on all closed ball in  $X$  since Lemma 4.1. Then there exists  $\exists x_1 \in \overline{B}_0$  and  $\exists n_1 \in E$  such that  $\|T_{n_1}x_1\| > 1$ . Since  $T_{n_1}$  is continuous and so is the norm, we have  $\|T_{n_1}x\| > 1$  on  $\overline{B}_1 = \overline{B}_{r_1}(x_1) \subset \overline{B}_0$ . Hence there exists  $\exists x_2 \in \overline{B}_1$  and  $n_2 > n_1, \exists n_2 \in E, \|T_{n_2}x_2\| > 2$ . Continuing in this way we obtain a sequence  $(x_k)$  and  $(\overline{B}_k)$  of closed balls such that  $x_k \in \overline{B}_k$  and  $\overline{B}_0 \supset \overline{B}_1 \supset \dots \supset \overline{B}_k \supset \dots$  and  $\|T_{n_k}x\| > k$  for  $n_k \in E$  and on the  $\overline{B}_k$ . This yields for  $n_1 < n_2 < \dots$  and  $\bar{x} \in X$ ,  $\|T_{n_k}\bar{x}\| > k$ , where  $n_k \in E$  for  $k = 1, 2, \dots$ . Hence we see that the sequence  $(T_n\bar{x})$  is not statistically bounded.

**Theorem 4.5.** Let  $T_k \in B(X, Y)$ , where  $X$  is a Banach space and  $Y$  a normed space. If  $(T_k)$  is strongly statistical operator convergent with limit  $T$ , then  $T \in B(X, Y)$ .

**Proof:** From Theorem 4.4, the proof is trivial.

## 5. STATISTICAL CORE OF SEQUENCES OF OPERATORS

Let  $(T_k)$  be a sequence of operators. For any sequence  $T = (\|T_k x\|)$ , the statistical limit superior of  $T$  and statistical limit inferior of  $T$  are

$$st - \limsup T_x = \begin{cases} \sup E_{T,x} & , \text{if } E_{T,x} \neq \emptyset \\ -\infty & , \text{if } E_{T,x} = \emptyset \end{cases}$$

$$st - \liminf T_x = \begin{cases} \inf F_{T,x} & , \text{if } F_{T,x} \neq \emptyset \\ +\infty & , \text{if } F_{T,x} = \emptyset \end{cases}$$

where  $E_{T,x} = \{a_x \in \mathbf{R} / \delta(\{k : \|T_k x\| > a_x\}) \neq 0\}$  and  $F_{T,x} = \{b_x \in \mathbf{R} / \delta(\{k : \|T_k x\| < b_x\}) \neq 0\}$  for every  $x \in X$ .

For example, consider the sequence  $(T_k)$  of operators  $T_k : \ell^1 \rightarrow \ell^1$  defined by

$$T_k x = \begin{cases} (k, 0, \dots, 0, \dots), & k = n^2 & n = 1, 2, \dots \\ (2\xi_1, \dots, 2\xi_n, \dots), & k \neq n^2 & \text{and } k = 2n - 1, \quad n = 1, 2, \dots \\ (3\xi_1, \dots, 3\xi_n, \dots), & k \neq n^2 & \text{and } k = 2n, \quad n = 1, 2, \dots \end{cases}$$

where  $x = (\xi_1, \dots, \xi_n, \dots) \in \ell^1$ . The operator  $T_k$  is linear and bounded for every fixed  $k \in \mathbf{N}$ . It is easy to see that  $E_{T,x} = (-\infty, 3\|x\|)$  and  $F_{T,x} = (2\|x\|, +\infty)$  since

$$\|T_k x\| = \begin{cases} k, & k = n^2 & n = 1, 2, \dots \\ 2\|x\|, & k \neq n^2 & \text{and } k = 2n - 1, \quad n = 1, 2, \dots \\ 3\|x\|, & k \neq n^2 & \text{and } k = 2n, \quad n = 1, 2, \dots \end{cases}$$

Clearly,  $\|T_k x\|$  is unbounded but it is statistical bounded. For this sequence,  $st - \limsup T_x = 3\|x\|$  and  $st - \liminf T_x = 2\|x\|$ . Furthermore, we can easily obtain from [4] that for any bounded sequence  $T = (\|T_k x\|)$  (i.e.  $\sup_{k \in \mathbf{N}} \|T_k x\| < \infty$ )

for every  $x \in X$ )

$st - \limsup T_x = \beta_x \iff$  for any  $\varepsilon > 0$  and for every  $x \in X$ ,  $\delta(\{k : \|T_k x\| > \beta_x - \varepsilon\}) \neq 0$  and  $\delta(\{k : \|T_k x\| > \beta_x + \varepsilon\}) = 0$ ; and  $st - \liminf T_x = \alpha_x \iff$  for any  $\varepsilon > 0$  and for every  $x \in X$ ,  $\delta(\{k : \|T_k x\| < \alpha_x + \varepsilon\}) \neq 0$  and  $\delta(\{k : \|T_k x\| < \alpha_x - \varepsilon\}) = 0$ .

We can define the statistical core of bounded linear operators as follows:

**Definition 5.1.** For any statistical bounded operator sequence  $T = (\|T_k x\|)$ , the statistical core of  $T$  is the closed interval  $[st - \liminf T_x, st - \limsup T_x]$ .

If  $(T_k)$  is not statistical bounded, the statistical core of  $T$  is defined by either  $(-\infty, st - \limsup T_x], [st - \liminf T_x, \infty)$  or  $(-\infty, \infty)$ .

Let  $X$  and  $Y$  be two nonempty subset of the spaces of complex sequences. Let  $A = (a_{nk})(n, k = 1, 2, \dots)$  be an infinite matrix. We write  $Ax = (A_n(x))$  if  $A_n(x) = \sum_{k=1}^{\infty} a_{nk}x_k$  converges for each  $n \in \mathbb{N}$ . Thus, we say that the matrix  $A$  defines a matrix transformation from  $X$  into  $Y$ .  $A$  is called regular if  $x \in c$  implies  $Ax \in c$  and preserves the limit, where  $c$  is convergent sequences space.

In [4] Fridy and Orhan prove necessary and sufficient conditions for which the inequalities

$$\limsup Ax \leq st - \limsup x$$

and

$$st - \liminf x \leq \liminf Ax$$

for every  $x \in \ell_{\infty}$ .

Now, Similarly, we will give these results for sequences of bounded linear operators.

Let  $A = (a_{nk})$  be an infinite summability matrix. For a given sequence of bounded linear operators  $\Gamma = (T_n)$ , the sums

$$AT_n x = \sum_{k=1}^{\infty} a_{nk} T_k x$$

are called the  $A$ -transform of the  $\Gamma$  provided the series converges for each  $k \in \mathbb{N}$  and for all  $x \in X$  and denoted by  $A\Gamma = (AT_n)$ .

**Lemma 5.1:** Let  $(T_n)$  be a sequence of uniformly bounded operators ( i.e. there is a positive number  $M$  such that  $\|T_n x\| < M$  for all  $x \in X$  and for all  $n \in \mathbb{N}$  )  $T_n \in B(X, Y)$ . Suppose the matrix  $A$  satisfies  $\sup_n \sum_{k=1}^{\infty} |a_{nk}| < \infty$  then

$$\limsup(\sup_{x \in X} \|AT_n x\|) \leq st - \limsup(\sup_{x \in X} \|T_n x\|) \quad (2)$$

if and only if

i)  $A$  is regular matrix and  $\lim_n \sum_{k \in E} |a_{nk}| = 0$  for  $E \in \mathbb{N}$  such that  $\delta(E) = 0$ ;

ii)  $\lim_n \sum_{k=1}^{\infty} |a_{nk}| = 1$ .

**Proof: (Necessity)** Assume that for any uniformly bounded sequence  $(\|T_n x\|)$  on  $Y$ , the matrix  $A = (a_{nk})$  satisfies condition (2). Let  $a_n = \sup_{x \in X} \|T_n x\|$ . Then

$y = (a_n)$  is a positive real number sequence and  $y \in \ell_{\infty}$ . Since  $\sup_n \sum_{k=1}^{\infty} |a_{nk}| < \infty$ ,

we obtain that

$$\begin{aligned} \sup_{n \in \mathbb{N}} (\sup_{x \in X} \|AT_n x\|) &\leq \sup_{n \in \mathbb{N}} (\sup_{x \in X} \sum_k |a_{nk}| \|T_k x\|) \\ &\leq \sup_{k \in \mathbb{N}} a_k (\sup_{n \in \mathbb{N}} \sum_k |a_{nk}|) < \infty, \end{aligned}$$

i.e.,  $(\sup_{x \in X} \|AT_n x\|) \in \ell_\infty$ . Thus, using Lemma in [4], conditions (i) and (ii) can be easily proved.

**(Sufficiency)** Assume now that (i) and (ii) holds. Let  $(\|T_n x\|)$  be any uniformly bounded sequence on  $Y$ . Then there is a positive number  $K$  such that  $\sup_{n \in \mathbb{N}} \|T_n x\| = K$  for all  $x \in X$ . Since  $\sup_n \sum_{k=1}^{\infty} |a_{nk}| < \infty$ , we have

$$\left\| \sum_k a_{nk} T_k x \right\| \leq \sum_k |a_{nk}| \|T_k x\| < K \sum_k |a_{nk}| < \infty.$$

Hence  $(AT_n x) \in \ell_\infty$ . Now, let  $\beta_x = st - \limsup (\sup_{x \in X} \|T_n x\|)$ . Then we have  $E = \{k : \sup_{x \in X} \|T_k x\| > \beta_x + \varepsilon\}$  for a given  $\varepsilon > 0$  and  $\delta(E) = 0$ . Hence it is clear that  $\sup_{x \in X} \|T_k x\| \leq \beta_x + \varepsilon$  for  $k \notin E$ . Now we can write

$$\begin{aligned} \left\| \sum_k a_{nk} T_k x \right\| &= \left\| \sum_k (|a_{nk}| \|T_k x\| + a_{nk} T_k x) 2^{-1} - \sum_k (|a_{nk}| \|T_k x\| - a_{nk} T_k x) 2^{-1} \right\| \\ &\leq \left\| \sum_k a_{nk} T_k x \right\| + \sum_k (|a_{nk}| - a_{nk}) \|T_k x\| \\ &\leq K \sum_{k \in E} |a_{nk}| + \sum_{k \notin E} |a_{nk}| (\sup_{x \in X} \|T_k x\|) + K \sum_k (|a_{nk}| - a_{nk}) \\ &\leq K \sum_{k \in E} |a_{nk}| + (\beta_x + \varepsilon) \sum_{k \notin E} |a_{nk}| + K \sum_k (|a_{nk}| - a_{nk}). \end{aligned}$$

Then we conclude that

$$\sup_{x \in X} \left\| \sum_k a_{nk} T_k x \right\| \leq K \sum_{k \in E} |a_{nk}| + (\beta_x + \varepsilon) \sum_{k \notin E} |a_{nk}| + K \sum_k (|a_{nk}| - a_{nk}).$$

Using the (i) and (ii), we have

$$\limsup_{x \in X} (\sup_{x \in X} \|AT_n x\|) \leq \beta_x + \varepsilon.$$

Since  $\varepsilon$  is arbitrary, this completes the proof. It is clear that one can prove a similar way in the following lemma.

**Lemma 5.2.** Let  $(T_k)$  be a sequence of uniformly bounded operators  $T_k \in B(X, Y)$ .

Suppose the matrix  $A$  satisfies  $\sup_n \sum_{k=1}^{\infty} |a_{nk}| < \infty$  then

$$st - \liminf (\inf_{x \in X} \|T_k x\|) \leq \liminf (\inf_{x \in X} \|AT_k x\|)$$

if and only if

i)  $A$  is regular matrix and  $\lim_n \sum_{k \in E} |a_{nk}| = 0$  for  $E \subset \mathbb{N}$  such that  $\delta(E) = 0$ ;

ii)  $\lim_n \sum_{k=1}^{\infty} |a_{nk}| = 1$ .

From Lemma 5.1 and 5.2, we give the following theorem.

**Theorem 5.1:** Let  $(T_k)$  be a sequence of uniformly bounded operators  $T_k \in B(X, Y)$ . Suppose the matrix  $A$  satisfies  $\sup_n \sum_{k=1}^{\infty} |a_{nk}| < \infty$  then

$$\limsup(\sup_{x \in X} \|AT_n x\|) \leq st - \limsup(\sup_{x \in X} \|T_n x\|)$$

and

$$st - \liminf(\inf_{x \in X} \|T_k x\|) \leq \liminf(\inf_{x \in X} \|AT_k x\|)$$

if and only if conditions (i) and (ii) in Lemma 5.2 are satisfied.

### References

- [1] *J. Connor* : Two valued measures and summability. Analysis 10 (1990), 373-385.
- [2] *H. Fast* : Sur la convergence statistique. Colloq. Math. 2 (1951), 241-244.
- [3] *J. A. Fridy* : On statistical convergence. Analysis 5 (1985), 301-313.
- [4] *J. A. Fridy and C. Orhan* : Statistical limit superior and limit inferior. Proc. Amer. Math. Soc. 125 (1997), 3625-3631.
- [5] *A. (Türkmenoğlu) Gökhan and M. Güngör* : On pointwise statistical convergence. Indian J. Pure appl. Math. 33(9) (2002), 1379-1384.
- [6] *T. Salat* : On statistically convergent sequences of real numbers. Math. Slovaca 30 (1980), 139-150.
- [7] *I. J. Schoenberg* : The integrability of certain functions and related summability methods. Amer. Math. Monthly 66 (1959), 361-375.
- [8] *B. C. Tripathy* : On statistically convergent sequences. Bull. Cal. Math. Soc. 90 (1998), 259-262

# THE FLOW OF A LIQUID WITH CAVITATION

Ivan Straškraba, Emil Vitásek\*

June 25, 2009

Institute of Mathematics of the Academy of Sciences of the Czech Republic,  
Prague, Czech Republic  
e-mail addresses: strask@math.cas.cz, vitas@math.cas.cz

## Abstract

The aim of this paper is to find, in a closed form, special solutions of equations describing a one-dimensional non-stationary flow of a liquid containing dissolved gas. The special feature of these solutions is that despite the fact that they do not satisfy all initial and boundary conditions, they describe a physical characteristic qualitatively analogous to that described by the original equations. Thus these special solutions may prove useful means for judging the reliability of the original mathematical model of the problem.

*Mathematics Subject Classification* (2000): 34A05

*Keywords*: Compressible fluid, Navier-Stokes equations, asymptotic behavior

## 1 Introduction

In the paper we analyze the mathematical model of the flow of a column of a real liquid.

It is known that cavitations play an important role not only in the theory of fluids but may be even more significant in the engineering and technological practice. Let us mention the cavities and bubbles which appear at suction compartments of pumps, in turbines, or in hydraulic machinery. Monitoring of possible separation of the gas from liquids is important since if it appears, then there is a danger of failure when enormous forces are loaded to devices which serve, for example, in the building industry. Therefore, a dynamic model of a two-phase flow has been derived under physically realistic assumptions. This has been done in the former Institute for Construction of Machines in Běchovice,

---

\*The research was supported by the Academy of Sciences of the Czech Republic, Institutional Research plan No. AV0Z10190503, and partially supported by the grant of the Grant Agency of the Czech Republic No. 201/08/0315



Czech Republic (see [1]). To preserve the author's and the institute's rights, we do not present here the derivation of the model and refer the interested readers to Professor Jan Šklíba at the Technical University of Liberec, Czech Republic (jan.skliba@tul.cz). The purpose of this paper is to partially mathematically analyze the model to be defined below. Our analysis is based on a special a priori assumed form of solutions to describe certain features of the flow in question. We balance this limitation by providing solutions in a closed form. In spite of this restriction we believe that our approach is useful in the engineering practice, based on the authors' experience in application of similar mathematical analysis to other problems.

## 2 The formulation of the problem

The equations of the flow of a real liquid of the length  $l$  is possible to write in the form (see for instance [1]):

$$w_t + \rho_0^{-1} p_x + f(w) = 0, \quad (2.1)$$

$$p_t + \rho_0 c^2(p, \gamma) w_x = 0, \quad (2.2)$$

$$\gamma_t + w \gamma_x = g(\gamma, p), \quad x \in (0, l), \quad t \in (0, T), \quad (T > 0), \quad (2.3)$$

$$w(x, 0) = w_0(x), \quad (2.4)$$

$$p(x, 0) = p_0(x), \quad (2.5)$$

$$\gamma(x, 0) = \gamma_0(x), \quad x \in [0, l], \quad (2.6)$$

$$C(p(0, t), \gamma(0, t)) p_t(0, t) + Q_V(p(0, t), H(t)) - S_0 w(0, t) + \varphi \dot{H}(t) = 0, \quad (2.7)$$

$$w(l, t) = h(t), \quad (2.8)$$

$$\ddot{H}(t) + \Phi(t, H(t), \dot{H}(t), p(0, t), p_t(0, t)) = 0, \quad t \in [0, T], \quad (2.9)$$

$$H(0) = H_0, \quad \dot{H}(0) = H_1. \quad (2.10)$$

The quantities occurring in (2.1)–(2.10) have the following meaning:

$w = w(x, t)$	the velocity of the liquid in the point $x$ and in the time $t$ ,
$p(x, t)$	the pressure,
$\gamma = \gamma(x, t)$	the mass of the freed air in the unit volume of the liquid,
$\rho_0$	the density of the liquid,
$c = c(p, \gamma)$	the sound velocity in the liquid and in the liquid containing the air, respectively (given function of $p, \gamma$ ),
$f = f(w)$	the coefficient of the resistance (the friction of the liquid on the wall of the duct), an odd function,
$g = g(\gamma, p)$	$= \begin{cases} K_u((\bar{\gamma} - \gamma)/K_H - p), & \text{if } (\bar{\gamma} - \gamma)/K_H \geq p, \\ K_r((\bar{\gamma} - \gamma)/K_H - p) & \text{if } (\bar{\gamma} - \gamma)/K_H < p, \end{cases}$
$K_u, K_r$	the constants characterizing the proportionality of the velocity of loosening, and dissolution on the pressure gradient, respectively,
$K_H$	the coefficient of absorption,

*The flow of a liquid with cavitation*

$\bar{\gamma}$	the total mass of the air in the unit volume,
$w_0, p_0, \gamma_0$	the initial distribution of the velocity, the mass, and the pressure of the loosened air in unit volume, respectively,
$C = C(p, \gamma)$	the hydraulic capacity (the given function of $p, \gamma$ ),
$H$	the throw of the valve,
$Q_V = Q_V(p, H)$	the flow through the valve (the given function of $p, H$ ),
$S_0$	the cross-section of the duct,
$\varphi$	the acting facing of the valve,
$h$	the flow rate caused by the hydrogenerator at the end of the duct,
$H_0, H_1$	the initial position, and the velocity of the valve, respectively.

In what follows, we assume that all given functions are sufficiently smooth, and the solution will be sought smooth as well, i.e., continuously differentiable.

The special solutions of our problem will be the functions  $w, p, \gamma$  satisfying equations (2.1), (2.2) and (2.3). The special feature of these solutions is that despite the fact that they do not satisfy all initial and boundary conditions, they express physical characteristics qualitatively analogous to those described by (2.1)–(2.10).

### 3 Stationary solution

Three functions  $w = w(x)$ ,  $p = p(x)$ ,  $\gamma = \gamma(x)$  depending only on the length coordinate  $x$  of the tube, and satisfying equations (2.1) to (2.3) are understood as stationary solution. These equations written for functions independent of the variable  $t$  form a simple system of three ordinary differential equations ([2])

$$\rho_0^{-1} p' + f(w) = 0, \quad (3.1)$$

$$\rho_0 c^2(p, \gamma) w' = 0, \quad (3.2)$$

$$w\gamma' = g(\gamma, p), \quad x \in (0, l), \quad (3.3)$$

where  $p' = dp/dx$  etc. Analogously as in [1] the function  $c(p, \gamma)$  will be assumed in the form

$$c(p, \gamma) = \frac{c_1 p^2}{c_2 p^2 + \gamma + c_3}, \quad (3.4)$$

where  $c_i > 0$ ,  $i = 1, 2, 3$  are constants. The physical principles suggest the condition  $c(p, \gamma) > 0$ . Since the trivial solution with  $p = 0$  is not interesting, the equation (3.2) gives us  $w' = 0$  and from here we have

$$w = w_0 = \text{constant}. \quad (3.5)$$

Consequently, (3.1) implies

$$p(x) = p_0 - \rho_0 f(w_0)x, \quad (3.6)$$

and for the function  $\gamma$  we obtain the equation

$$\gamma' = \frac{1}{w_0} g(\gamma, p_0 - \rho_0 f(w_0)x). \quad (3.7)$$

The constants  $w_0$ ,  $p_0$  may be chosen arbitrarily. Also the integration of (3.7) gives an additional free integration constant. The heuristic considerations indicate that the stationary solution should be a limit of a non stationary solution for  $t \rightarrow \infty$ . In order to respect the boundary conditions partially at least we will require the stationary solution to satisfy the generalized limit of boundary conditions. More exactly, we impose the condition (compare with the (2.8))

$$w_0 = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t h(s) ds. \quad (3.8)$$

Naturally, we must suppose that the function  $h$  is such that this limit exists. The limit is requested in that sense since the delivery of the hydrogenerator may regularly oscillate about a certain value so that the limit  $\lim_{t \rightarrow \infty} h(t)$  need not exist. The number  $p_0$  in (3.6) is then determined from (2.7), i.e., from

$$Q_V(p_0, H_0) - S_0 w_0 = 0 \quad (3.9)$$

supposing that equation (3.9) is solvable with respect to  $p_0$ . It remains to determine the function  $\gamma$  from (3.7). It is obvious that the sign of  $w_0$  indicates the direction of the flow. If the hydrogenerator is supposed to be at the point  $x = l$  it is logical that  $w_0 < 0$  and then it is necessary to describe  $\gamma(l) = \gamma_0$  – the mass of loosening air in the unit volume of incoming liquid. Since the function  $g$  in the right-hand term of (3.7) is given by different formulas for loosening and dissolution of the air we must distinguish between the following two cases:

$$g(\gamma, p) = \begin{cases} K_u \left( \frac{\bar{\gamma} - \gamma}{K_H} - p \right), & \text{if } \frac{\bar{\gamma} - \gamma}{K_H} \geq p \\ K_r \left( \frac{\bar{\gamma} - \gamma}{K_H} - p \right), & \text{if } \frac{\bar{\gamma} - \gamma}{K_H} < p. \end{cases} \quad (3.10)$$

Let us define the functions

$$\varphi(x) = \bar{\gamma} - K_H[p_0 - \rho_0 f(w_0)x], \quad (3.11)$$

and

$$K(\xi) = \begin{cases} \frac{K_u}{w_0 K_H} \equiv -K_1, & \text{if } \xi \geq 0 \\ \frac{K_r}{w_0 K_H} \equiv -K_2, & \text{if } \xi < 0, K_i > 0, i = 1, 2. \end{cases} \quad (3.12)$$

Then it is possible to rewrite equation (3.7) in the form

$$\gamma' = (\varphi(x) - \gamma) \cdot K(\varphi(x) - \gamma), \quad (3.13)$$

*The flow of a liquid with cavitation*

or, if we put

$$y(x) = \varphi(x) - \gamma(x), \quad (3.14)$$

$$y' + yK(y) = \varphi'(x). \quad (3.15)$$

According to (3.11) it is

$$\varphi'(x) = K_H \rho_0 f(w_0) \equiv \varphi_0 = \text{const.} \quad (3.16)$$

If  $(\bar{\gamma} - \gamma_0)/K_H \geq p(l) = p_0 - \rho_0 f(w_0)l$ , then  $y(l) \geq 0$ , and the continuity of the solution  $y$  of (3.15) implies  $K(y) = K_u/w_0 K_H = -K_1 = \text{const.}$  for  $x < l$  sufficiently close to  $l$ . Consequently, equation (3.15) is linear on this interval and we have

$$\begin{aligned} y(x) &= \exp(K_1(x-l))y(l) + \int_l^x \exp(K_1(x-\xi))\varphi_0 d\xi \\ &= \exp(K_1(x-l))y(l) + \frac{\exp(K_1(x-l)) - 1}{K_1} \varphi_0. \end{aligned} \quad (3.17)$$

Since  $w_0 < 0$  and the function  $f(w)$  describing the friction is necessarily odd and positive for positive  $w$ 's, from (3.16) and (3.17), it follows that  $y(x) > \exp(K_1(x-l))y(l) \geq 0$  in the *whole* interval  $[0, l]$ . But this fact means that the function  $y(x)$  is defined for *all*  $x \in [0, l]$  by formula (3.17). Using (3.17), (3.14) and (3.11) we then obtain for the function  $\gamma(x)$  the formula

$$\begin{aligned} \gamma(x) &= \varphi(x) - y(x) = \bar{\gamma} - K_H[p_0 - \rho_0 f(w_0)x] \\ &\quad - \exp(-K_1(l-x))(\bar{\gamma} - K_H[p_0 - \rho_0 f(w_0)l] + \gamma_0) \\ &\quad - \frac{1 - \exp(-K_1(l-x))}{K_1} K_H \rho_0 f(w_0) \end{aligned} \quad (3.18)$$

for  $x \in [0, l]$ , and  $K_1 = -K_u/w_0 K_H$ .

On the other hand, let  $(\bar{\gamma} - \gamma_0)/K_H < p(l) = p_0 - \rho_0 f(w_0)l$ , i.e.,  $y(l) < 0$ . Then it follows – again from the continuity – that  $K(y) = -K_2$  in (3.15) for  $x \in (x^*, l]$  where  $x^*$  is the upper bound of numbers in the interval  $(-\infty, l)$  satisfying  $y(x^*) = 0$ . On the interval  $(x^*, l]$ , the solution  $y(x)$  is given by formula

$$y(x) = \exp(K_2(x-l))y(l) + \frac{\exp(K_2(x-l)) - 1}{K_2} \varphi_0. \quad (3.19)$$

If  $x^* \leq 0$  then the solution of (3.15) is given by (3.19) in the whole interval  $[0, l]$ . From the condition  $y(x^*) = 0$  and from (3.19) we obtain the unique point

$$x^* = l + \frac{1}{K_2} \ln \left( \frac{\varphi_0}{\varphi_0 + K_2 y(l)} \right). \quad (3.20)$$

If  $x^* > 0$  which means the assumption

$$\gamma_0 < \bar{\gamma} - K_H[p_0 - \rho_0 f(w_0)l] + \frac{K_H^2 \rho_0 f(w_0)w_0}{K_r} \left( \exp\left(-\frac{K_r l}{w_0 K_H}\right) - 1 \right), \quad (3.21)$$

as we obtain after elementary calculation and substitution from (3.16), (3.14) and (3.12), then  $y(x)$  is given by (3.19) only for  $x \in (x^*, l]$ . In the interval  $[0, x^*]$  we then easily obtain that

$$\begin{aligned} y(x) &= \exp(K_1(x - x^*))y(x^*) + \int_{x^*}^x \exp(K_1(x - \xi))\varphi_0 d\xi \\ &= \frac{K_H^2 \rho_0 w_0 f(w_0)}{K_u} \left(1 - \exp\left(\frac{K_u}{w_0 K_H}(x^* - x)\right)\right), \end{aligned} \quad (3.22)$$

( $y(x^*) = 0$ !), and for  $x^*$  it is necessary to substitute from (3.20). If we substitute from (3.14), (3.11) in the formulae (3.19) and (3.12), respectively we obtain the formula for  $\gamma(x)$  analogously as in (3.18).

The stationary problem is completely solved.

## 4 Oscillatory solution

Oscillatory solutions are called such solutions of equations (2.1) to (2.3) which do not depend on the space variable  $x$ . Thus,  $w = w(t)$ ,  $p = p(t)$ ,  $\gamma = \gamma(t)$ . In the special case, the system (2.1) to (2.3) may be written as follows:

$$\dot{w} + f(w) = 0, \quad (4.1)$$

$$\dot{p} = 0 \quad (4.2)$$

$$\dot{\gamma} = g(\gamma, p), \quad t > 0. \quad (4.3)$$

We see that the equations are practically separated since if we compute

$$p(t) = p_0 = \text{const.}, \quad (4.4)$$

from (4.2) we have two separate equations, (4.1) and

$$\dot{\gamma} = g(\gamma, p_0). \quad (4.5)$$

The solvability of equation (4.1) for  $t \in (0, \infty)$  is guaranteed by the assumption

$$m = \inf_{w \in \mathbb{R}} f'(w) > -\infty. \quad (4.6)$$

( $f'(w)$  is continuous). The proof of this assertion is an elementary consequence of the theory of ordinary differential equations. Assumption (4.6) is practically always fulfilled. If, for example,  $f(w) = k|w|w$  ( $k > 0$  constant), then  $f'(w) = 2k|w| \geq 0 = m$ . As far as equation (4.5) is concerned the global existence of the solution is guaranteed by the inequality  $|g(\gamma_1, p_0) - g(\gamma_2, p_0)| \leq \max\{K_u/K_H, K_r/K_H\}|\gamma_1 - \gamma_2|$ ,  $\gamma_1, \gamma_2 \in \mathbb{R}$  which follows from (3.10). This condition means nothing else than that the function  $g$  is globally Lipschitzian with respect to the variable  $\gamma$ . Naturally, it is possible, for  $w$  and  $\gamma$ , to prescribe the initial conditions

$$w(0) = w_0, \quad \gamma(0) = \gamma_0. \quad (4.7)$$

*The flow of a liquid with cavitation*

This is a possibility how to satisfy, at least partially, the initial conditions (2.4) to (2.6), naturally only with constant functions  $w_0(x) = w_0$ ,  $p_0(x) = p_0$ ,  $\gamma_0(x) = \gamma_0$ . On the other hand, boundary conditions (2.7), (2.8) will be never satisfied by this type of solution.

Equations (4.1), (4.5), (4.7) can be solved numerically by well known approximate methods for solving ordinary differential equations. The solution in a closed form can be obtained only for a particular choice of the function  $f$ . Hence, let us investigate the case

$$f(w) = k|w|w. \quad (4.8)$$

If  $w_0 > 0$  then we have, due to (4.1), (4.8),  $\int_{w_0}^w dw/kw^2 = -t$  and from here

$$w(t) = \frac{w_0}{1 + kw_0 t}. \quad (4.9)$$

If  $w_0 < 0$  we obtain analogously

$$w(t) = \frac{w_0}{1 - kw_0 t}. \quad (4.10)$$

The formulae (4.9), (4.10) may be, for both cases, joined in one

$$w(t) = \frac{w_0}{1 + k|w_0|t}. \quad (4.11)$$

Equation (4.5) may be now rewritten in the form

$$\dot{y} + yK(y) = 0, \quad (4.12)$$

where  $y(t) = \bar{\gamma} - p_0 K_H - \gamma(t)$ ,

$$K(\xi) = \begin{cases} \frac{K_u}{K_H}, & \text{if } \xi \geq 0 \\ \frac{K_r}{K_H}, & \text{if } \xi < 0. \end{cases} \quad (4.13)$$

If  $y_0 \equiv \bar{\gamma} - p_0 K_H - \gamma_0 \geq 0$ , the solution of (4.12) is the function  $y(t) = \exp(-(K_u/K_H)t)y_0$ . If  $y_0 < 0$  then  $y(t) = \exp(-(K_r/K_H)t)y_0$ . Using (4.13) we have from here

$$\gamma(t) = \bar{\gamma} - p_0 K_H - \exp\left(-\frac{K_u}{K_H}t\right), \quad \text{if } \gamma_0 \leq \bar{\gamma} - p_0 K_H \quad (4.14)$$

and

$$\gamma(t) = \bar{\gamma} - p_0 K_H - \exp\left(-\frac{K_r}{K_H}t\right), \quad \text{if } \gamma_0 > \bar{\gamma} - p_0 K_H. \quad (4.15)$$

The oscillatory solution is given by formulae (4.4), (4.11), (4.14) and (4.15).

Here the oscillatory solution do not respect fully its appellation. Namely, the functions (4.14), (4.15) stabilize for  $t \rightarrow \infty$  to the steady state value  $\bar{\gamma} - p_0 K_H$  without any overswing.

## 5 Combined solutions

By combined solution we mean in this context such solution of equations (2.1) to (2.3) which is neither stationary - nor oscillatory, and for which at least one of the functions  $w$ ,  $p$ ,  $\gamma$  depends only on  $x$  or  $t$ .

Consider first such solutions, for which  $w = w(t)$ ,  $p = p(x)$ . For this case the equations (2.1), (2.2), (2.3) have the form

$$\dot{w} + \frac{1}{\rho_0} p' + f(w) = 0, \quad (5.1)$$

$$\gamma_t + w\gamma_x = g(\gamma, p). \quad (5.2)$$

Equation (2.2) is obviously satisfied identically. From (5.1) it follows

$$p'(x) = \text{const}. \quad (5.3)$$

Instead of (2.7), (2.8), choose for  $p$  the boundary conditions

$$p(0) = p_0, \quad p(l) = p_1, \quad (5.4)$$

which imitate the gradient of the pressure caused by the hydrogenerator. Then

$$p(x) = p_0 + \frac{p_1 - p_0}{l} x, \quad (5.5)$$

as it follows from (5.3). Consequently, (5.1) gives

$$\dot{w} + f(w) = \frac{p_0 - p_1}{\rho_0 l} \quad (5.6)$$

for the function  $w$ . If we suppose (4.6) and supply the initial condition

$$w(0) = w_0 \quad (5.7)$$

we know that there exists a global solution  $w(t)$  of problem (5.6), (5.7). If we know such solution we can find the function  $\gamma$  from the equation

$$\gamma_t + w(t)\gamma_x = g\left(\gamma, p_0 + \frac{p_1 - p_0}{l} x\right) \quad (5.8)$$

by the method of characteristics. Naturally it is necessary to add the corresponding initial and boundary conditions. The initial condition may be given in the quite general form

$$\gamma(x, 0) = \gamma_0(x). \quad (5.9)$$

If it is  $w(t) > 0$ , we must prescribe

$$\gamma(0, t) = \gamma^0(t), \quad (5.10)$$

*The flow of a liquid with cavitation*

i.e., the concentration of the air in the liquid flowing into the tube from the end  $x = 0$ . If  $w(t) < 0$  it is necessary to prescribe

$$\gamma(l, t) = \gamma^1(t), \quad (5.11)$$

for the liquid flowing into the tube from the end  $x = l$ . Boundary conditions (5.10), (5.11) may change one for other during the time but never can be prescribed both together since  $w$  is independent of  $x$ . Hence, the problem given by (5.8), (5.9), and (5.10) and (5.11), respectively must be solved after such time-intervals, on which the function  $w(t)$  does not change the sign. In order to avoid the complicated testing let us choose a particular situation which is also applicable in practice.

Suppose that

$$\begin{aligned} p_1 > p_0, \quad w < 0, \quad \frac{p_0 - p_1}{\rho_0 l} - f(w_0) < 0, \\ f(-\xi) = -f(\xi), \quad f'(\xi) \geq 0, \quad \xi \in \mathbb{R}. \end{aligned} \quad (5.12)$$

Physically, these assumptions mean

- (i) the pressure on the left-hand side of the tube is greater than on the right-hand side;
- (ii) the liquid flows from the right-hand side to the left-hand side at the beginning;
- (iii) the difference of the pressures is not yet weighted by the force of the resistance of the duct at the beginning;
- (iv) the force of the resistance always acts against the direction of the motion of the fluid and it does not abate with increasing velocity.

Assumptions (5.12) and equations (5.6), (5.7) imply  $\dot{w} < 0$ . Thus  $w(t)$  is a decreasing function for increasing  $t$  either for any  $t > 0$  or there exists a  $t^* > 0$  such that  $f(w(t^*)) = (p_0 - p_1)/\rho_0 l$  and then  $w(t) = w(t^*)$  for  $t \geq t^*$ . The reason is that the function  $w(t) = w(t^*)$  is a solution of equation (5.6) on the interval  $(t^*, \infty)$  with the initial condition  $w(t^*)$  and that the equation (5.6) has the unique solution for the given initial condition. Hence  $w(t) < 0$  for all  $t \geq 0$  and we have to solve the problem (5.8), (5.9), (5.11) for determining  $\gamma$ .

Let us apply the method of characteristics. Let  $x \in [0, l]$ ,  $t > 0$  be arbitrary. Put

$$\mathcal{X}(\tau; x, t) = x + \int_t^\tau w(s) ds. \quad (5.13)$$

Then  $\mathcal{X}_\tau(\tau; x, t) = w(\tau)$ . If we put, moreover,

$$\phi(\tau) = \gamma(\mathcal{X}(\tau; x, t), \tau). \quad (5.14)$$



*I. Straškraba, et al.*

we have

$$\begin{aligned}\dot{\phi}(\tau) &= \gamma_x \mathcal{X}_\tau + \gamma_t = \gamma_t + w\gamma_x \\ &= g(\phi(\tau), p_0 + \frac{p_1 - p_0}{l} \mathcal{X}(\tau; x, t)).\end{aligned}\quad (5.15)$$

Thus if  $x$  and  $t$  are given we obtained a differential equation for the function  $\phi(\tau)$  given by (5.14). If we find  $\phi(\tau)$  for  $\tau \in [0, t]$ , then

$$\gamma(x, t) = \phi(t) \quad (5.16)$$

as it follows from (5.14), (5.13). The initial condition for the function  $\phi(\tau)$  is given either by the formula

$$\phi(0) = \gamma(\mathcal{X}(0; x, t), 0) = \gamma_0(x - \int_0^t w(s)ds) \quad (5.17)$$

if the characteristic  $\xi = \mathcal{X}(\tau; x, t)$  falls on the axis  $\tau = 0$  in the interval  $0 \leq \xi \leq l$  or by the formula

$$\phi(\tau^*) = \gamma(l, \tau^*) = \gamma^1(\tau^*) \quad (5.18)$$

where  $\mathcal{X}(\tau^*; x, t) = l$  if this characteristic falls on the axis  $\xi = l$  at some point  $\tau = \tau^* > 0$ . The value of  $\tau^*$  is to be computed from an implicit equation

$$x - \int_{\tau^*}^t w(s)ds = \mathcal{X}(\tau^*; x, t). \quad (5.19)$$

Put

$$y(\tau) = \varphi(\tau; x, t) - \phi(\tau), \quad (5.20)$$

where

$$\varphi(\tau; x, t) = \bar{\gamma} - K_H \left( p_0 + \frac{p_1 - p_0}{l} \left( x + \int_t^\tau w(s)ds \right) \right) \quad (5.21)$$

and

$$K(\xi) = \begin{cases} \frac{K_u}{K_H}, & \text{if } \xi \geq 0 \\ \frac{K_r}{K_H}, & \text{if } \xi < 0. \end{cases} \quad (5.22)$$

Then (5.15) can be written in the form  $\dot{y} + yK(y) = \varphi_\tau$  and it is, according to (5.21) and (5.12),  $\varphi_\tau = K_H((p_0 - p_1)/l)w(\tau) > 0$ . Solve the equation

$$\dot{y} + yK(y) = K_H \frac{p_0 - p_1}{l} w(\tau) \quad (5.23)$$

*The flow of a liquid with cavitation*

with the initial condition

$$\begin{aligned} y(0) &= \varphi(0; x, t) - \phi(0) \\ &= \bar{\gamma} - K_H \left( p_0 + \frac{p_1 - p_0}{l} \left( x + \int_t^0 w(s) ds \right) \right) - \gamma_0 \left( x - \int_0^t w(s) ds \right) \equiv y_0 \end{aligned} \quad (5.24)$$

which corresponds to the situation when the characteristic  $\xi = \mathcal{X}(\tau; x, t)$  falls on the axis  $\xi$ . If it is now  $y_0 \geq 0$ , then the solution of problem (5.23), (5.24) is the function

$$y(\tau) = \exp\left(-\frac{K_u}{K_H}\tau\right)y_0 + K_H \frac{p_0 - p_1}{l} \int_0^\tau \exp\left(-\frac{K_u}{K_H}(\tau - s)\right)w(s)ds \quad (5.25)$$

for  $\tau \in [0, t]$ . If  $y_0 < 0$ , then the solution of problem (5.23), (5.24) is the function

$$y(\tau) = \exp\left(-\frac{K_r}{K_H}\tau\right)y_0 + K_H \frac{p_0 - p_1}{l} \int_0^\tau \exp\left(-\frac{K_r}{K_H}(\tau - s)\right)w(s)ds \quad (5.26)$$

as far as  $y(\tau) < 0$ . If  $\tau_1 \equiv \inf\{\tau; 0 < \tau \leq t, y(\tau) = 0\} < t$ , then it is necessary to prolong the solution (5.26) onto the interval  $[\tau_1, t]$  by the formula

$$y(\tau) = K_H \frac{p_0 - p_1}{l} \int_{\tau_1}^\tau \exp\left(-\frac{K_u}{K_H}(\tau - s)\right)w(s)ds. \quad (5.27)$$

When the characteristic  $\xi = \mathcal{X}(\tau; x, t)$  falls on the axis  $\xi = l$  we proceed completely analogously; as an initial condition we use

$$\begin{aligned} y(\tau^*) &= \varphi(\tau^*; x, t) - \phi(\tau^*) \\ &= \bar{\gamma} - K_H \left( p_0 + \frac{p_1 - p_0}{l} \left( x + \int_t^{\tau^*} w(s) ds \right) \right) - \gamma^1(\tau^*) \equiv y_1 \end{aligned} \quad (5.28)$$

and integrate the equation over the interval  $(\tau^*, t]$ . At the same time the value  $\tau^* = \tau^*(x, t)$  will be computed from (5.19). Again, it is necessary to distinguish whether  $y_1 \geq 0$  or  $y_1 < 0$ . In the first case we obtain

$$\begin{aligned} y(\tau) &= \exp\left(-\frac{K_u}{K_H}(\tau - \tau^*)\right)y_1 \\ &+ K_H \frac{p_0 - p_1}{l} \int_{\tau^*}^\tau \exp\left(-\frac{K_u}{K_H}(\tau - s)\right)w(s)ds, \quad \tau \in [\tau^*, t]. \end{aligned} \quad (5.29)$$

In the second case it is necessary, moreover, to distinguish whether  $y(\tau_1) = 0$  for some (smallest one)  $\tau_1 \in (\tau^*, t)$  or  $y(\tau) < 0$  for all  $\tau \in (\tau^*, t)$ . In the latter case we have

$$y(\tau) = \exp\left(-\frac{K_r}{K_H}(\tau - \tau^*)\right)y_1 + K_H \frac{p_0 - p_1}{l} \int_{\tau^*}^\tau \exp\left(-\frac{K_r}{K_H}(\tau - s)\right)w(s)ds \quad (5.30)$$

*I. Straškraba, et al.*

for all  $\tau \in [\tau^*, t]$ . In the opposite case the formula (5.30) is valid only for  $\tau \in [\tau^*, \tau_1]$  and for  $\tau \in (\tau_1, t)$ , we must, moreover,  $y(\tau)$  prolong using the formula

$$y(\tau) = K_H \frac{p_0 - p_1}{l} \int_{\tau_1}^{\tau} \exp\left(-\frac{K_u}{K_H}(\tau - s)\right) w(s) ds, \quad \tau \in (\tau_1, t). \quad (5.31)$$

Finally, the value of  $\gamma(x, t)$  is found from formula (5.16), where  $\phi(t) = \varphi(t; x, t) - y(t)$ ,  $\varphi$  is given by (5.21),  $y(t)$  by formulae (5.25) to (5.28), the values  $y_0$  and  $y_1$  by (5.24) and (5.28), respectively, and  $\tau^*$  by equation (5.19).

Hence, if

$$\gamma_0\left(x - \int_0^t w(s) ds\right) < \bar{\gamma} - K_H\left(p_0 + \frac{p_1 - p_0}{l}\left(x + \int_t^0 w(s) ds\right)\right), \quad (5.32)$$

then, according to (5.26), (5.20) and (5.25), we have

$$\begin{aligned} \gamma(x, t) &= \phi(t) = \varphi(t; x, t) - y(t) \\ &= \bar{\gamma} - K_H\left(p_0 + \frac{p_1 - p_0}{l}\left(x + \int_t^0 w(s) ds\right)\right) \\ &\quad - \exp\left(-\frac{K_u}{K_H}t\right) \left(\bar{\gamma} - K_H\left(p_0 + \frac{p_1 - p_0}{l}\left(x + \int_t^0 w(s) ds\right)\right)\right) \\ &\quad - \gamma_0\left(x - \int_0^t w(s) ds\right) - K_H \frac{p_0 - p_1}{l} \int_0^t \exp\left(-\frac{K_u}{K_H}(t - s)\right) w(s) ds. \end{aligned} \quad (5.33)$$

In other cases we proceed analogously. We will not introduce the resulting formulae since the procedure consists in fact only in the routine substitution even though formally complicated. It is clear that we may not be able to compute  $w(t)$  in a closed form for general function  $f$ . Consequently also the formulae for  $\gamma(x, t)$  will not be explicit in general.

On the other hand let us introduce the concrete formulae for the physically interesting special case, namely, that  $f(w) = k|w|w$ . Then equation (5.6) has the form

$$\dot{w} = \frac{p_0 - p_1}{\rho_0 l} + kw^2 \quad (5.34)$$

since  $w < 0$  as it follows from our preceding investigations and thus  $k|w|w = -kw^2$ . Moreover it is  $p_1 > p_0$ . Put

$$z(t) = \left(\frac{k\rho_0 l}{p_1 - p_0}\right)^{1/2} w(t), \quad t > 0, \quad z(0) = z_0 = \left(\frac{k\rho_0 l}{p_1 - p_0}\right)^{1/2} w_0. \quad (5.35)$$

Then (5.34) can be written in the form

$$\frac{d}{dt} \left( \log \left( \frac{z - 1}{z + 1} \right) \right) = 2 \left( \frac{k(p_1 - p_0)}{\rho_0 l} \right)^{1/2}. \quad (5.36)$$

*The flow of a liquid with cavitation*

Integrating (5.36) over the interval  $(0, t)$  we get

$$\log\left(\frac{z-1}{z+1} \frac{z_0+1}{z_0-1}\right) = \alpha t, \quad (5.37)$$

where

$$\alpha = 2\left(\frac{k(p_1 - p_0)}{\rho_0 l}\right)^{1/2}. \quad (5.38)$$

Applying to (5.37) the function exp we obtain

$$\frac{z-1}{z+1} = \frac{z_0-1}{z_0+1} e^{\alpha t}. \quad (5.39)$$

After straightforward but rather lengthy calculation we arrive at the formula

$$w(t) = \left(\frac{p_1 - p_0}{k\rho_0 l}\right)^{1/2} z(t) = \left(\frac{p_1 - p_0}{k\rho_0 l}\right)^{1/2} \frac{\alpha w_0 + 2 + (\alpha w_0 - 2)e^{\alpha t}}{\alpha w_0 + 2 + (\alpha w_0 - 2)e^{\alpha t}}, \quad (5.40)$$

where  $\alpha$  is given by (5.38) and  $k$  is the constant of friction in (4.8).

Notice, that

$$\lim_{t \rightarrow \infty} w(t) = -\left(\frac{p_1 - p_0}{k\rho_0 l}\right)^{1/2}$$

which corresponds to the equilibrium state. Since the limit speed of the fluid is negative, the fluid flows from the end  $x = l$  in direction to  $x = 0$ . This in accordance with the physical idea that the fluid flows from the place of higher pressure to the region, where the pressure is lower.

All three special solutions which we have introduced have a common deficiency. They are not influenced by the dependence of the sound speed  $c = c(p, \gamma)$  on the pressure and the concentration of the air since the member  $c^2(p, \gamma)w_x$  in equation (2.2) in all cases vanishes. From that reason it seems reasonable to drop the member  $f(w)$  by putting it equal to zero and to investigate the solution of the type  $w = w_1 x + w_0$ ,  $p = p(t)$ ,  $\gamma = \gamma(t)$  where  $w_0$ ,  $w_1$  are constants. Then equation (2.1) (with  $f \equiv 0$ ) is satisfied identically and from (2.2), (2.3) we obtain the system of two ordinary differential equations

$$\begin{aligned} \dot{p}(t) &= \rho_0 c^2(p(t), \gamma(t)), \\ \dot{\gamma}(t) &= g(p(t), \gamma(t)), \quad t > 0 \end{aligned}$$

with the initial conditions

$$\begin{aligned} p(0) &= p_0 = \text{const.} \\ \gamma(0) &= \gamma_0 = \text{const.} \end{aligned}$$

The analytic investigation and the numerical solution of this simple system can bring the tentative idea of the influence of the dependence of  $c = c(p, \gamma)$  on the behaviour of the system and thus to obtain comparative characteristic for the solution of the general problem and for the correspondence with the physical concept of the behaviour of the system.

*I. Straškraba, et al.*

## References

- [1] J. Šklíba, I. Straškraba, M. Štengl, *Extended mathematical model of safety hydraulic circuit*. Report SVÚSS Běchovice, Czech Republic registered as: SVÚSS 88-03022, December 1988.
- [2] E.A.Coddington, N.Levinson, *Theory of Differential Equations*. McGraw Hill, New York, 1955.

---

**Instructions to Contributors**  
**Journal of Concrete and Applicable Mathematics**  
 A quarterly international publication of Eudoxus Press, LLC, of TN.

**Editor in Chief: George Anastassiou**  
 Department of Mathematical Sciences  
 University of Memphis  
 Memphis, TN 38152-3240, U.S.A.

**1. Manuscripts hard copies in triplicate, and in English, should be submitted to the Editor-in-Chief:**

**Prof. George A. Anastassiou**  
 Department of Mathematical Sciences  
 The University of Memphis  
 Memphis, TN 38152, USA.  
 Tel. 901.678.3144  
 e-mail: [ganastss@memphis.edu](mailto:ganastss@memphis.edu)

Authors may want to recommend an associate editor the most related to the submission to possibly handle it.

Also authors may want to submit a list of six possible referees, to be used in case we cannot find related referees by ourselves.

**2. Manuscripts should be typed using any of TEX, LaTeX, AMS-TEX, or AMS-LaTeX and according to EUDOXUS PRESS, LLC. LATEX STYLE FILE. (Click [HERE](#) to save a copy of the style file.) They should be carefully prepared in all respects. Submitted copies should be brightly printed (not dot-matrix), double spaced, in ten point type size, on one side high quality paper 8(1/2)x11 inch. Manuscripts should have generous margins on all sides and should not exceed 24 pages.**

**3. Submission is a representation that the manuscript has not been published previously in this or any other similar form and is not currently under consideration for publication elsewhere. A statement transferring from the authors (or their employers, if they hold the copyright) to Eudoxus Press, LLC, will be required before the manuscript can be accepted for publication. The Editor-in-Chief will supply the necessary forms for this transfer. Such a written transfer of copyright, which previously was assumed to be implicit in the act of submitting a manuscript, is necessary under the U.S. Copyright Law in order for the publisher to carry through the dissemination of research results and reviews as widely and effectively as possible.**

**4. The paper starts with the title of the article, author's name(s) (no titles or degrees), author's affiliation(s) and e-mail addresses. The affiliation should comprise the department, institution (usually university or company), city, state (and/or nation) and mail code.**

**The following items, 5 and 6, should be on page no. 1 of the paper.**

**5. An abstract is to be provided, preferably no longer than 150 words.**

**6. A list of 5 key words is to be provided directly below the abstract. Key words should express the precise content of the manuscript, as they are used for indexing purposes.**

**The main body of the paper should begin on page no. 1, if possible.**

**7. All sections should be numbered with Arabic numerals (such as: 1. INTRODUCTION) .**

**Subsections should be identified with section and subsection numbers (such as 6.1. Second-Value Subheading).**

**If applicable, an independent single-number system (one for each category) should be used to label all theorems, lemmas, propositions, corollaries, definitions, remarks, examples, etc. The label (such as Lemma 7) should be typed with paragraph indentation, followed by a period and the lemma itself.**

**8. Mathematical notation must be typeset. Equations should be numbered consecutively with Arabic numerals in parentheses placed flush right, and should be thusly referred to in the text [such as Eqs.(2) and (5)]. The running title must be placed at the top of even numbered pages and the first author's name, et al., must be placed at the top of the odd numbered pages.**

**9. Illustrations (photographs, drawings, diagrams, and charts) are to be numbered in one consecutive series of Arabic numerals. The captions for illustrations should be typed double space. All illustrations, charts, tables, etc., must be embedded in the body of the manuscript in proper, final, print position. In particular, manuscript, source, and PDF file version must be at camera ready stage for publication or they cannot be considered.**

**Tables are to be numbered (with Roman numerals) and referred to by number in the text. Center the title above the table, and type explanatory footnotes (indicated by superscript lowercase letters) below the table.**

**10. List references alphabetically at the end of the paper and number them consecutively. Each must be cited in the text by the appropriate Arabic numeral in square brackets on the baseline.**

**References should include (in the following order):  
initials of first and middle name, last name of author(s)  
title of article,**

name of publication, volume number, inclusive pages, and year of publication.

Authors should follow these examples:

### **Journal Article**

1. H.H.Gonska, Degree of simultaneous approximation of bivariate functions by Gordon operators, (journal name in italics) *J. Approx. Theory*, 62,170-191(1990).

### **Book**

2. G.G.Lorentz, (title of book in italics) *Bernstein Polynomials* (2nd ed.), Chelsea, New York, 1986.

### **Contribution to a Book**

3. M.K.Khan, Approximation properties of beta operators, in (title of book in italics) *Progress in Approximation Theory* (P.Nevai and A.Pinkus, eds.), Academic Press, New York, 1991, pp.483-495.

11. All acknowledgements (including those for a grant and financial support) should occur in one paragraph that directly precedes the References section.

12. Footnotes should be avoided. When their use is absolutely necessary, footnotes should be numbered consecutively using Arabic numerals and should be typed at the bottom of the page to which they refer. Place a line above the footnote, so that it is set off from the text. Use the appropriate superscript numeral for citation in the text.

13. After each revision is made please again submit three hard copies of the revised manuscript, including in the final one. And after a manuscript has been accepted for publication and with all revisions incorporated, manuscripts, including the TEX/LaTeX source file and the PDF file, are to be submitted to the Editor's Office on a personal-computer disk, 3.5 inch size. Label the disk with clearly written identifying information and properly ship, such as:

Your name, title of article, kind of computer used, kind of software and version number, disk format and files names of article, as well as abbreviated journal name.

Package the disk in a disk mailer or protective cardboard. Make sure contents of disks are identical with the ones of final hard copies submitted!

Note: The Editor's Office cannot accept the disk without the accompanying matching hard copies of manuscript. No e-mail final submissions are allowed! The disk submission must be used.

14. Effective 1 Nov. 2009 for current journal page charges, contact the Editor in Chief. Upon acceptance of the paper an invoice will be sent to the contact author. The fee payment will be due one month from the invoice date. The article will proceed to publication only after the fee is paid. The charges are to be sent, by money order or certified check, in US dollars, payable to Eudoxus Press, LLC, to the address shown on



the Eudoxus [homepage](#).

No galleys will be sent and the contact author will receive one(1) electronic copy of the journal issue in which the article appears.

15. This journal will consider for publication only papers that contain proofs for their listed results.

# **TABLE OF CONTENTS, JOURNAL OF CONCRETE AND APPLICABLE MATHEMATICS, VOL. 8, NO. 4, 2010**

<b>On the numerical solution of nonlinear delay differential equations, S. Karimi Vanani, A. Aminataei,.....</b>	<b>568</b>
<b>Inclusion Theorems for Absolute Matrix Summability Methods, W. T. Sulaiman,....</b>	<b>577</b>
<b>Integral Inequalities Concerning Triple Integrals, W. T. Sulaiman,.....</b>	<b>585</b>
<b>On some inequalities for Concave Functions, W. T. Sulaiman,.....</b>	<b>594</b>
<b>Recurrence relation with binomial coefficient, George Grossman, Aklilu Zeleke, Xinyun Zhu,.....</b>	<b>602</b>
<b>On the q-extension of Genocchi polynomials, C. S. Ryoo,.....</b>	<b>616</b>
<b>On Best Simultaneous Approximation in Semi Metric Spaces, H. K. Pathak and Satyaj Tiwari, .....</b>	<b>623</b>
<b>On best uniform approximation of periodic functions by trigonometric polynomials, Michael I. Ganzburg,.....</b>	<b>631</b>
<b>Some convergence theorems for a class of generalized <math>\Phi</math>-hemicontractive mappings, Chang He Xiang, Zhe Chen, Ke Quan Zhao,.....</b>	<b>638</b>
<b>Iterative algorithms for a countable family of nonexpansive mappings, Yisheng Song, Xiao Liu,.....</b>	<b>645</b>
<b>Statistical Convergence and Statistical Core of Sequences of Bounded Linear Operators, A. Gökhan,.....</b>	<b>656</b>
<b>The flow of a liquid with cavitation, Ivan Straskraba, Emil Vitasek,.....</b>	<b>668</b>